12-2016

Estimation of Annual Average Daily Traffic (AADT) and Missing Hourly Volume Using Artificial Intelligence

Sababa Islam
Clemson University

ESTIMATION OF ANNUAL AVERAGE DAILY TRAFFIC (AADT) AND MISSING HOURLY VOLUME USING ARTIFICIAL INTELLIGENCE

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Civil Engineering

---

by
Sababa Islam
December 2016

Accepted by:
Mashrur Chowdhury, Committee Chair
Wayne Sarasua
Feng Luo

ABSTRACT

Annual Average Daily Traffic (AADT) is one of the most important traffic parameters used in transportation planning and engineering analysis. Moreover, each state Department of Transportation (DOT) must report the AADT data to the Federal Highway Administration (FHWA) annually as part of the Highway Performance Monitoring System (HPMS) reporting requirements. For this reason, state DOTs continually collect AADT data via permanent count stations and short-term counts. In South Carolina, only interstates and primary routes are equipped with permanent count stations. For the majority of the secondary routes, AADT data are estimated based on short-term counts or are simply guesstimated based on their functional classifications. In this study the use of Artificial Neural Network (ANN) and Support Vector Regression (SVR) were applied to estimate AADT from short-term counts. These estimated AADTs were compared to the traditional factor method used by South Carolina Department of Transportation (SCDOT) and also to the Ordinary Least-square Regression method. The comparison between ANN and SVR revealed that SVR functions better than ANN in AADT estimation for different functional classes of roadways. A second comparison was conducted between SVR and the traditional factor method. A comparative analysis revealed that SVR performed better that the traditional factor method. Similarly, the comparison between SVR and regression analysis, for the principal arterials, revealed no significant difference in the actual AADT and AADTs estimated through SVR. However, it did show a significant difference between the actual AADT and AADT estimated through regression analysis.

One of the primary challenges of accurate measurement of AADT is having reliable, complete, and accurate traffic data. Previous literature indicated that often the transportation agencies reported the problem of missing hourly volume from the permanent traffic count stations. These studies reported that the percentage of missing traffic data vary between 10% to 60%. In an effort to address this issue, most of the state departments of transportation either discard or impute the missing data. SCDOT imputes the missing hourly volume using the historical average of the last 3 months' data from the same day and hour. This method of data imputation could often be erroneous. In order to develop an accurate estimation of missing hourly volume from the permanent count stations, this study applied two Artificial Intelligence Paradigms, Artificial Neural Network (ANN) and Support Vector Regression (SVR) for predicting hourly missing data. Data imputation models were developed for Urban Principal Arterial (Interstate), Rural Principal Arterial (Interstate), and Urban Principal Arterials-other functional class. Each of these functional classes were divided into different ANN and SVR models based on the on different combination of input features. This study indicated that for each functional class, SVR outperformed ANN. The SVR model performance was later compared with current SCDOT's imputation practice, which revealed that SVR model is more accurate in estimating missing values compared to the imputation method by SCDOT.

DEDICATION

I would like to dedicate this thesis to my grandparents, my parents and my youngest uncle for their unconditional love and support. My youngest uncle, who himself is a renowned Civil Engineer, dreamt of me being a Civil Engineer since the day I was born, and has greatly contributed to my passion for this field.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation and gratitude to my advisor, Dr. Mashrur Chowdhury for his continuous guidance, inspiration, and support throughout my journey as a master's student. He is the one who motivated me to endeavor challenges that I never imagined I could accomplish. I can't thank him enough for believing in me and for involving me in different research activities and projects.

I would also like to thank Dr. Wayne Sarasua and Dr. Feng Luo for serving as my thesis committee members. Thank you for reviewing my thesis, and providing valuable insights about the research.

I would like to specially acknowledge my better half, Sakib Mahmud Khan, who has done everything for my ease and comfort during my entire journey as an MS student. It's he who took care of everything when I was busy with my thesis. I simply could not be able to earn this degree without him having beside me.

I would like to thank the South Carolina Department of Transportation (SCDOT) for providing me with the data that were necessary for my research.

I would like to extend my deep appreciation to Dr. Kakan Dey for his continuous effort to improve the research quality and for always being there whenever I needed him. I would like to specially thank Md. Mizanur Rahman to help me out in times when I was in need of suggestions for my research. I sincerely thank Joshua Mitchell and McKenzie Keehan for being the best colleagues that I can ask for. I appreciate both of them for reviewing my work. I am grateful to Dr. Katalin Beck, for reviewing my thesis when I was desperately in need of a technical writing expert. I also acknowledge the help from the

Clemson University Writing Center while writing this thesis. Moreover, I will like to recognize Md Mhafuzul Islam and Md Zadid khan for helping me to improve my thesis defense.

I cordially thank my parent-in-laws, my sister and my niece and everyone else from my family for being the support system for me during any critical times. They have always been my inspiration to reach my goals.

I appreciate the staff members from the Glenn Department of Civil Engineering, and Kristin Baker in particular for extending her help with any types of administrative work. Finally, I would like to express my wholehearted gratitude to the Bangladeshi Community in Clemson for making Clemson my home and giving me the warmth of a family.

Table of Content

List of Tables

List of Figures

CHAPTER ONE

INTRODUCTION

1.1 Background and Motivation

Annual Average Daily Traffic (AADT) is one of the most important parameters in transportation engineering. It is calculated by adding the total vehicle volume of a highway for a year divided by 365 days. It is one of the most important traffic measures used in any transportation related projects (i.e. roadway design, transportation planning, traffic safety analysis, highway investment decision making, highway maintenance, air quality compliance study and travel demand modeling). It is also an important input variable for safety analysis and is used in Safety Analyst software and the Highway Safety Manual (Harwood, 2004). Moreover, as a part of the traffic monitoring program, every state department of transportation has to report the AADT on federal aid highways to FHWA annually (TMG, 2016). Thus, the accuracy of AADT estimation is critical to any transportation problems that uses AADT as an input parameter. However, to develop an accurate method of estimating AADT is one of the biggest challenges in transportation engineering keeping in mind the lack of enough funding.

An accurate means of measuring AADT for a road segment involves installing permanent traffic count stations or Automatic Traffic Recorders (ATRs). An ATR collects traffic data 24 hours a day and 365 days a year using traditional inductive loops, microwave radar sensors, magnetic counters, and piezoelectric sensors. However, installation of the permanent count stations using the traditional technologies at thousands of traffic count stations throughout a given network to estimate AADT data is hardly economical (Atluri, et al., 2009); therefore, ATRs are installed only at a limited number of locations and short term traffic counts (i.e., 24/48-hour) are performed at most of the other locations where an AADT estimation is required. These short term counts are expanded using some calibration factors to calculate AADT, which is known as the Factor Method. The data collection frequencies at short term count stations are inconsistent among states. While short-term counts are performed annually in some states, others span a few years (Sharma et al., 1999).

Traditional AADT estimation method entails the use of expansion factors (seasonal, daily, monthly, growth and axle adjustment factors) to the volume collected from the short-term traffic count stations. This method of AADT estimation involves 1) calculating the expansion/adjustment factors using the data from the continuous traffic count stations, and 2) applying the calculated factors to the roadway locations with short-term counts to estimate AADT (Garber and Hoel, 2014). In order to develop reliable adjustment factors, permanent and short terms count stations are grouped together based on the geographical locations and the functional class of roadway. After grouping, permanent count station data are used to develop the average adjustment factors, and short-term count locations within the same group is used to estimate AADT by applying these factors. This method of AADT estimation at short term count station is quite ambiguous since there are no defined guidelines or established standards regarding the method of assigning the expansion factors from ATR to the short-term traffic count stations (Sharma et al., 1999). Moreover, the relatively small number of ATRs in the lower functional class of roadways makes it challenging for the development of accurate expansion factors for large number of short term count stations on local roads. Which creates the need for more permanent count stations in the lower functional classes. Researchers have used several alternative methods for estimating AADT, which include regression analysis, regression analysis using centrality and roadway characteristic variables, travel demand modelling, machine learning techniques, image processing to circumvent the limitations of the traditional AADT estimation methods (Sharma et al., 1999 and Keehan et al., 2017).

The key for estimating accurate AADT is the availability of reliable, accurate and complete traffic data. These traffic data are not only used to calculate AADT but also to estimate Design Hourly Volume (DHS), average travel speed, and to forecast the future traffic conditions. Specific traffic data, such as volumes of traffic, speed data, occupancy rates are used for designing the traffic control system. Despite calculating traffic parameters and designing traffic control systems, transportation agencies are now more inclined to use real time traffic data for transportation network optimization with increasing travel demand. As mentioned earlier transportation agencies usually collect traffic data from permanent count stations continuously for 365 days a year, it is challenging to obtain accurate and complete data without any missing and inaccurate values due to several factors, such as hardware or software malfunctioning on data collection equipment and technology or loss of data packages during transmission from roadside ATRs to traffic data processing centers (Qu et al., 2009). Multiple previous studies have identified the extent of missing data at ATRs. A study by Zong et al. indicated that on an average, ATRs have more than 50% of values missing, based on data collected from Alberta, Minnesota, and

Saskatchewan ATRs (Zhong et al., 2004). Similarly, the percentage of missing data from some loop detectors in the California performance measurement system (PeMS) is higher than 10% (Performance Measurement System, 2016). South Carolina Department of Transportation (SCDOT) is not an exception. Due to the missing data of the permanent count stations (i.e., ATRs), traffic parameters (i.e. AADT) often have to be estimated based on incomplete data, which can lead to estimation inaccuracies.

In order to overcome this limitation, transportation agencies often impute these missing hourly volume. It is mentioned in the AASHTO guidelines that if the missing traffic data is not extensive with respect to the entire data collected from a particular location and if the missing data is randomly scattered throughout the year, traffic agencies may impute hourly volume (Vandervalk-Ostrander, 2009). However, it is also mentioned in the guideline that, there should be a threshold (not more than 50% of the data) for the percentage of missing data and if missing data exceeds that threshold, agencies should not use that data for developing traffic statistics (Vandervalk-Ostrander, 2009). Although the transportation agencies impute missing traffic data, the Traffic Monitoring Guide (TMG) and AASHTO guidelines have particularly mentioned the importance of "Truth-in-Data", and it is recommended that if state DOTs adjust/impute missing data they should maintain record of the data adjustment procedure (TMG, 2016 and Vandervalk-Ostrander, 2009).

1.2 Research Objectives

The specific objectives of this research are as follows:

1. Develop AADT estimation models using machine learning techniques for different functional classes of roadways in South Carolina;
2. Compare the AADT estimated by machine learning techniques and traditional factor method used by SCDOT.
3. Develop missing hourly volume imputation models for different ATR locations using machine learning techniques.
4. Compare the missing hourly volume imputed by models using machine learning techniques and the historical average method used by SCDOT.

1.3 Organization of the Thesis

The thesis is comprised of five chapters. Chapter 1 of the thesis consists of the research background and motivation for this study, followed by the research objectives of the thesis. Chapter 2 summarizes the review of different AADT estimation methods and missing hourly volume imputation methods. Chapter 3 presents the method describing how the Artificial Intelligence (AI) based models were developed for estimating AADT and imputing missing hourly volume traffic data. Chapter 4 summarizes the results of the AADT estimated using different ANN and SVR based models in the study and comparison of the AADT estimated for the AI based models developed in the study with the factor based method currently used by SCDOT. This chapter also presents the results of hourly missing hourly volume imputation developed my machine learning techniques and compare the results with the historical average method used by SCDOT. Finally, Chapter 5 concludes the thesis with the important research finding and recommendations based on the results.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

This chapter is divided into two sections. Section 2.2 presents the review of different AADT estimation methods and their efficacies. The method includes:

- Traditional Factor method
- Regression analysis
- Machine learning techniques

Section 2.3 of this chapter presents the different missing hourly volume imputation methods that have been reviewed for this research. The following is a list of methods presented in this section:

- Interpolation-based Imputation Methods
- Statistical Learning-Based Imputation Methods
- Prediction-Based Imputation Methods

2.2 Different methods for AADT Estimation

This section summarizes the different AADT estimation methods that have been reviewed for this research.

2.2.1 Traditional Factor Method

Traditional factor method is the most widely adopted method for estimating AADT in USA. According to a survey conducted by a research project, it was found that among the 39 participating state DOTs 35 of them use factor method for estimating AADT from the short term traffic count stations (Islam et al., 2017). While the traffic monitoring guide and AASHTO have provided guidelines for estimating AADT using the factor methods, state DOTs usually improvise it according to their specific needs (TMG, 2001 and AASHTO, 1994). In this method, the short term traffic counts (24, 48 or 72 hours) taken at some strategic roadway locations are adjusted using different expansion factors. These factors include seasonal, axle adjustment factors and growth factors. The mathematical formulation of the AADT using the factor method is as follows

$$AADT_{gi} = ADT_{gi} \times AF_i \times SF_g \times GF_g$$

$AADT_{gi}$ = the annual average daily traffic at location i of factor group g,

$ADTh_i$ = the average daily (vehicle/axle) traffic at location i of factor group g,

AF_i = the applicable axle correction factor for location i (if needed),

SF_g = the applicable seasonal adjustment factor for group g, and

GF_g = the applicable annual growth factor for group g (if needed).

Permanent count stations data are used to develop these factors. The estimation of these factors is critical for calculating accurate estimate of AADT. Usually the ATRs are grouped and the factors developed from each ATR locations are averaged. The ATR stations are grouped based on roadway functional class, land use or geographic location in most of the time. The factors developed are than applied to an individual or to a group of short term traffic count stations. There are no defined guidelines on how to assign the factors to the short term traffic count stations which often leads to inaccurate estimation of AADT.

2.2.2 Regression Analysis

Regression analysis is one of the most popular methods for AADT estimation. Having incorporated demographic variables into the estimation model, Mohammad et al. found that county arterial mileage and county population were two significant quantitative independent variables (Mohammad et al. 1998). They also found that location and accessibility were two significant qualitative variables effecting the volume of traffic on the paved county roads. Roadways characteristics in AADT estimation in Florida were considered by Xia et al. (1999). GIS technology was used by Zhao and Chung (2001) to extract land-use and accessibility information to be used in regression models. However, few studies addressed modified version of the regression models. Geographically weighted regression (GWR) was applied by Zhao and Park (2004) to estimate regression parameters locally instead of globally. The comparison showed that GWR is more accurate than ordinary linear regression (OLR). Jiang et al. (2006) proposed to use a weighted average of i) growth factor method, which uses last years' data to predict AADT and ii) traffic count from current year's image. Kingan and Westhuis (2006) proposed a regression method that is more robust in estimating AADT than the ordinary least square method, since the ordinary least square method is vulnerable to outliers. Yang et al. (2011) studied variable selection and parameter estimation using different groups of variables. The variable selection by smoothly clipped absolute deviation penalty (SCAD) method can select significant variables and estimate regression coefficients simultaneously. Important variables can be selected using the smoothly clipped absolute deviation penalty (SCAD) method. Regression coefficients can also be estimated using this method simultaneously.

2.2.3 Machine Learning Techniques

For the last decades, machine learning has been gaining constant attention in the field of transportation engineering (Bhavser et al., 2007). Among the different algorithms, ANN has been used extensively in studying driver behaviors, maintenance of pavement, classification or detection of vehicles, analysis of traffic patterns and forecasting of traffic (Himanen et al. 1998). In addition, Sharma et al. used hourly volume factors as the predictor variable for estimating AADT. Here, they determined the effectiveness of two or more short-term traffic counts that were collected at different periods of the traffic counting season over the traditional method of AADT estimation. While they determined that the traditional method outperformed the ANN, the reason for this superior performance was the accurate grouping of the permanent and short-term count stations, which is rare in practical cases (Sharma et al., 1999). In their follow up study using hourly volume from 55 permanent count stations to inform ANN for AADT for lower volume roadways of Alberta, Canada, they also found that the traditional factor method to be superior (Sharma et al., 2001). However, they also found that because estimating AADT using ANN does not require grouping of the permanent count stations, there is no need to correctly assign short-term count stations to an ATR group. Therefore, in such a case ANN is recommended.

SVR being another form of machine learning techniques is one of the most common applications of SVM. This method uses a set of supervised learning methods and can be successfully applied for regression similar to the ANN. A study by Lin indicated that SVR has greater learning potential than ANN (Lin, 2004). However, limited research has been conducted using SVR in traffic data analysis (Vanajakshi and Laurence, 2004). The applications of SVR and SVM in the field of transportation engineering include; its use for travel time prediction, incident detection, real-time highway traffic condition assessment and development of decision support system for real-time traffic management (Ma et al., 2012; Ma et al., 2010; Ma et al. 2009, Chowdhury et al. 2006 and Bhavsar et al., 2007).

Vanajakshi and Laurence (2004) found that when training data was limited, SVR performed better then ANN for predicting short–term traffic. For the years between 1985 and 2004, Castro-Neto et al. (2009) used AADT values for urban and rural roads in 25 different counties in Tennessee for evaluating the performance of a modified version of SVR named SVR with Data-dependent Parameters (SVR-DP). An evaluation of the SVR-DP approach with the Ordinary OLS-regression methods and popular Holt Exponential Smoothing (Holt-ES) revealed that the SVR-DP outperformed both, although the Holt-ES also performed well for estimating AADT.

2.3 Different Methods for Imputing Missing Hourly Volume

In order to execute traffic management and traffic flow pattern predictions, a reasonable amount of traffic count data is necessary, both temporally and spatially. The technologies used for traffic data collection often produce missing or erroneous data. In an attempt to mitigate these missing data, a variety of data imputation methods have been developed. These methods have been divided into three main types: interpolation-based, statistical learning-based, and prediction-based. These methods are discussed in the following subsections.

2.3.1 Interpolation-based Imputation Methods

In the Interpolation-based methods missing data is imputed using a weighted average of known data that is either pattern neighboring or temporal-neighboring. For example, in a study by Zhong et al. (2004) developed, Autoregressive Integrated Moving Average (ARIMA), neural network and regression models. The study found that regression models that are genetically designed based on data from before and after the imputation performed better than other methods. The average errors of these models were lower than 1%. A time-delay neural network and locally-weighted regression model were developed by Zhong et al. based on genetic algorithm which had higher accuracy than the traditional imputation models. For the genetically designed neural network model and regression model the 95th percentile errors were below 6% and 2% respectively. Imputation accuracy of the models is influenced to some extent by the underlying traffic pattern, revealed by the study results based on sample traffic counts from different functional classes and trip pattern groups. However, it is clear that in most cases, genetically designed regression models can bound the 95th percentile errors to less than 5% (Zhong et al. 2004).

2.3.2 Statistical Learning-Based Imputation Methods

Statistical feature of traffic flow is used in the statistical learning-based methods. The method assumes a special probability distribution of the experiential data. Using this method missing data are imputed using the data that best fit the assumed probability distribution. Robust Principal Component Analysis (PCA) was applied by Qu et al. (2009) to filter the unusual traffic flow data that disturb the imputation process. In addition to this, the authors compared the performance of PPCA/Bayesian PCA-based imputation algorithms with different conventional methods (i.e. nearest/mean historical imputation methods and the local interpolation/regression methods). The results from the study revealed that, the PPCA based methods reduced the root-mean-square imputation error by at least 25% than the conventional methods.

In order to predict the freeway travel time, Van Lint (2005) developed a framework that exploits a recurrent neural network topology which is called state space neural network (SSNN). The SSNN is designed based on the layout of the freeway stretch of interest. This proposed SSNN combines the traffic related design with the generality of the neural network approaches. In this method simple imputation methods like spatial interpolation and exponential forecasts are used for imputing missing data. Results from the study revealed that, SSNN generated a MRE of 1.5% and a standard deviation of the relative error of 6.5% on the larger data set. However, on the smaller set, the errors increased within a reasonable range.

Asif et al. (2013) proposed methods that can construct a low-dimensional representation of large and diverse networks in the presence of missing historical and neighboring data to reconstruct data profiles for road segments, and impute missing values. They use Fixed Point Continuation with Approximate SVD (FPCA) and Canonical Polyadic (CP) decomposition for incomplete tensors to solve the problem of missing data. They concluded that FPCA and CP-WOPT can reconstruct traffic profiles with decent accuracy, even from very sparse data sets. The methods work well for expressway networks as well as large urban settings containing a diverse set of road segments.

2.3.3 Prediction-Based Imputation Methods

Two missing data imputation methods were developed by Nelwamondo (2010); 1. Expectation Maximization (EM) Algorithm and 2. A combination of auto-associative Neural Networks and Genetic Algorithm. These two types of methods performed differently based on the relationship among the independent variables. Results for the study revealed that, Expectation Maximization performs better when the input variables are either independent or minimally related to each other. However, the combination of auto-associative neural network and genetic algorithm performed well when there are some inherent non-linear relationships between some of the given variables.

In order to impute the holiday traffic, Liu et al. (2008) developed a K-nearest neighbor (K-NN). The k-NN method is a data-driven non-parametric regression method which is renowned for modeling unusual conditions. Regardless of the season that holidays are observed and how high or low the traffic volumes are, their observed minimum estimation errors (MinARE) were always near zero, and their MARE and median errors (E50) were generally in the range of 6-10%.

Regression models, Neural Network model that is designed with generic algorithm, the traditional factor method and Autoregressive Integrated Moving Average (ARIMA) models were used by Sharma et al. (2003) for missing hourly data imputation. They developed imputation models for different roadway functional classes and traffic pattern groups using the data from 6 permanent count stations. Moreover, they tested how the accuracy of imputation using these methods effect the estimation of AADT and DHV. Study results revealed that the AADT and DHV estimation models are higher for the traditional factor method. The study results also showed that among the different methods studied in this study, genetically designed neural network produced the least error in estimation AADT and DHV.

CHAPTER THREE

RESEARCH METHOD

3.1 Overview

The two major objectives of this research were to develop models to estimate AADT for the different functional classes of roadways in South Carolina, and to develop models for imputing missing hourly volume for the permanent traffic count stations. In order to develop models for estimating AADT, two Artificial Intelligence (AI) paradigms (i.e., Artificial Neural Network and Support Vector Regression) have been used. Following the development of the models, the results were evaluated and were compared with the traditional factor based AADT estimation method currently used by SCDOT and a traditional regression analysis method for different roadway functional classes.

To develop models for imputing missing hourly volume, two Artificial Intelligence paradigms have been used and the results were compared with the historical average method of missing data imputation currently used by SCDOT.

In this chapter the Artificial Intelligence paradigms that have been used in this study have been introduced. Each step of the method for developing the models for estimating AADT and imputing missing hourly volume is descried in greater depth.

3.2 AADT Estimation Using Machine Leaning Techniques

This section outlines the methods used in the AADT estimation model development using Artificial Intelligence. **Figure 3-1** illustrates five- phased method followed for developing of AADT estimation models using AI (Please see section 3.2.3 for detail information).

a) Urban Principal Arterial- Interstate and Expressways
b) Rural Principal Arterial- Interstate
c) Urban Principal Arterial – Other
d) Rural Principal Arterial- Other
e) Combination of All Functional Classes

Each of the phases are described in detail in the following section.

3.2.1 Phase 1: ATR Data Collection

The AADT estimation models were developed for different functional class of roadways in South Carolina using two types of data.

a) The hourly volume collected from all permanent count stations operating 365 days a year for the year 2011.
b) Census data collected from the census database to represent socio-economic characteristics of cities where permanent count stations are located.

SCDOT maintains a total number of 150 permanent count stations (i.e., ATR) on different functional classes with most on higher volume highways (**Figure 3-2)** and **Figure 3-3** shows a sample of the data reported in the website. For this research, hourly volume counts for all ATRs were collected for year 2011.

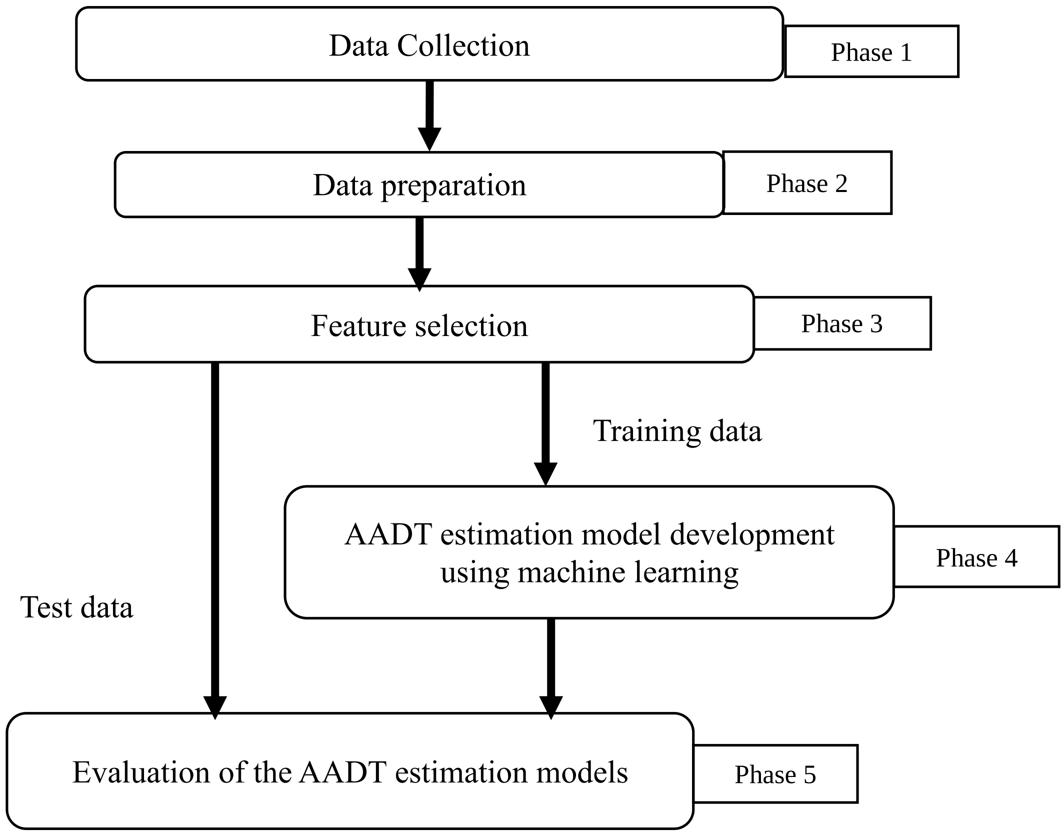

Figure 3-1 AADT Estimation Method

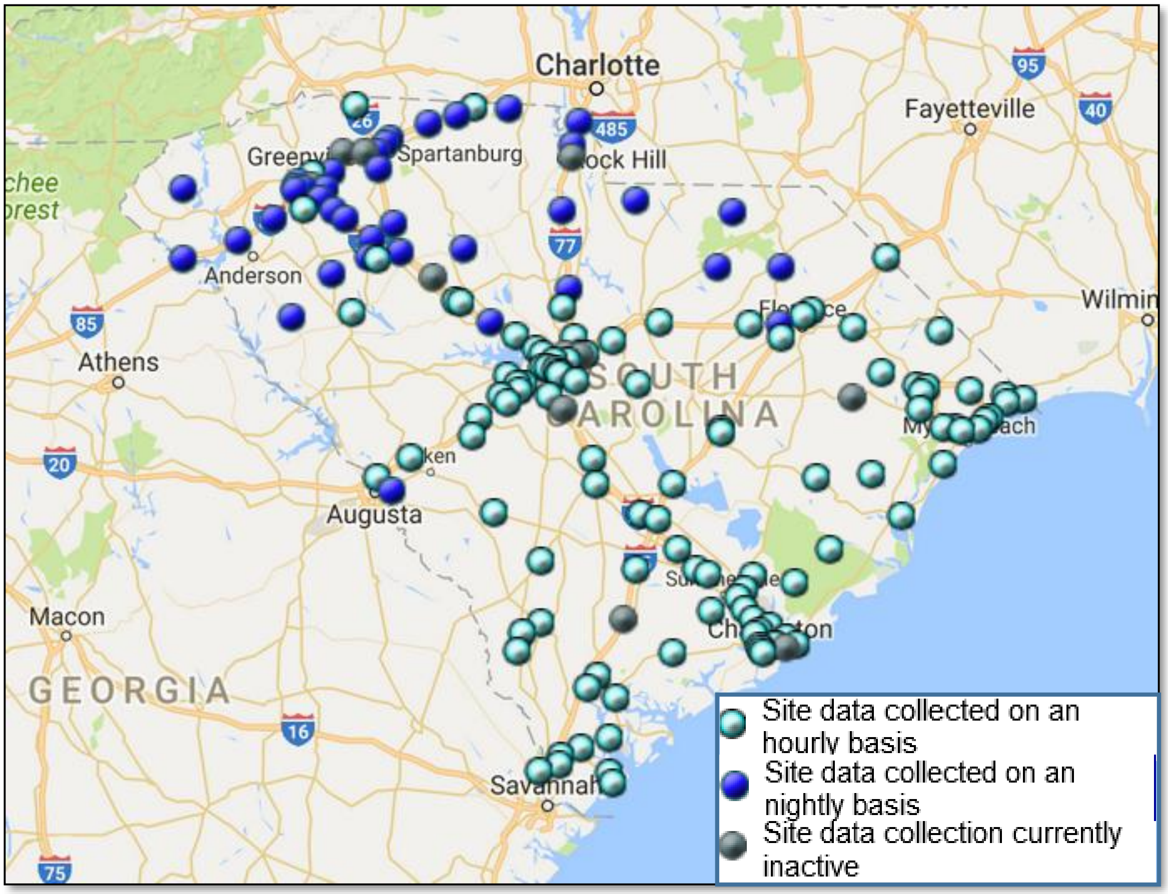

Figure 3-2 ATR Locations in South Carolina (**Source:** http://dbw.scdot.org/Poll5WebAppPublic/wfrm/wfrmHomePage.aspx)

Site: 0006-CHARLESTON for Monday, 01/03/2011

Site Location: US-17 Ravenel Br to SC-703 (MP 31.51 - MP 34.19)

Time	Northbound			Southbound		
	Vehicle Count		Average Speed (MPH)	Vehicle Count		Average Speed (MPH)
	Current	Historical		Current	Historical	
01:00	51	52	47	55	52	43
02:00	27	32	48	18	27	41
03:00	31	26	47	13	17	43
04:00	15	17	44	13	14	45
05:00	20	26	46	29	27	47
06:00	56	66	47	103	112	46
07:00	176	185	46	334	405	46
08:00	592	590	47	961	1168	47
09:00	747	846	48	1083	1285	47
10:00	616	691	48	694	785	46
11:00	557	633	49	576	703	46
12:00	640	751	49	650	707	45
13:00	770	847	50	714	772	45
14:00	749	824	50	825	811	45
15:00	876	891	49	776	859	46
16:00	929	1151	49	814	885	45
17:00	1264	1416	48	828	920	45
18:00	1376	1826	47	898	1038	45
19:00	743	942	48	597	706	44
20:00	424	526	48	367	437	44
21:00	242	353	49	244	366	44
22:00	189	266	47	263	277	44
23:00	139	164	46	152	207	43
24:00	91	107	45	89	99	42

Figure 3-3 Sample One-day Data for Station Table of Contents (**Source:** http://dbw.scdot.org/Poll5WebAppPublic/wfrm/wfrmHomePage.aspx)

Data is collected from the SCDOT website using an interactive web crawling model developed in Python 2.7.10 using a library called Selenium (Muthukadan, 2016). Selenium library is an Application program interface (API) on the object Web driver. Web driver works as a browser which can load a website and interacts with the different page elements. Web driver has the capability to fill forms and crawl through the web site like a human user and simulate mouse clicks (Web scraping 2016). **Figure 3-4** presents the data collection procedure from the SCDOT website for collecting data from the 134 ATR stations using selenium Web Driver. According to the Traffic Monitoring Guide, the presence of missing data in the permanent count stations can produce biased AADT (TMG, 2014). Therefore, the ATRs with more than six months missing data were not used for developing the models. Data were collected for the year of 2011 for the all the 134 ATRs. Hourly counts for a day was removed from the records if any hourly volume for that day was missing, caused by data collection equipment hardware or software malfunctions, or loss of data package during transmission in intelligent transportation systems (Qu et al. 2009).

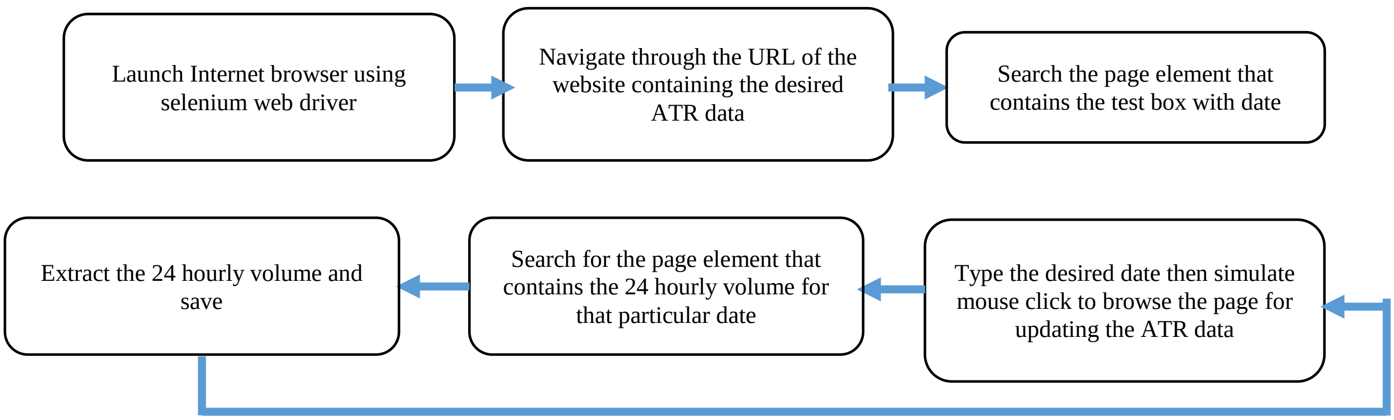

Figure 3-4 Data collection method

In addition, census data was collected considering land use around ATRs (Smith, 2016). The census data used for developing the models are

- Income
- Employment
- Percent below poverty
- Number of vehicles

- Urban or rural
- Number of housing units

All of these data were collected from the census database for the year 2011. In addition to these data, categorical features (Day of week and Month of Year) and another feature for number of lanes at each ATR were also used. A detailed description of the preparation of the data is discussed in next section 3.2.2.

3.2.2 Phase 2: Data Preparation

In order to develop and evaluate the AI models using machine learning techniques, two types of features were used

a) **input features:** hourly volume factors, socio economic data from census database, number of lane and categorical features (day of week, and hours of the day)

b) **target feature:** AADT factor which is a factor obtained by diving the actual AADT of an ATR station by the 24 hourly volume of a day.

Once the data are prepared the entire data set (i.e., one-year worth of hourly volume counts for all 117 ATRs) is separated into training and testing cases.

a. Training Data: This data is for developing the learning algorithm for predicting AADT. As a rule of thumb for developing the AI models 2/3 (two-third) of the data from the data set is used for training purpose (Mitchell 1998).

b. Test Data: This data is only used for testing the performance of the models developed using training data, and should be totally independent of the training data set. In this study, 1/3 (one third) of the data from the data set is used for testing purpose.

Following sections presents the detailed description about how the data were prepared for developing the models.

Input Features 1 to 24- Hourly Volume Factor Data Preparation: To develop the AADT estimation models 24 hourly volume factors were used. The formula for developing the hourly volume factor is expressed below:

Hourly volume factor for hour x =

$$\frac{Traffic\ volume\ for\ \ hour\ x\ (e.g.,\ traffic\ volume\ for\ 7AM-8AM\ on 1st\ \ monday\ of\ january, 2011)}{Sum\ of\ 24\ hourly\ volume\ of\ that\ day} \ldots\ldots\ldots (1)$$

Input Feature- Socio-economic Data Preparation: In addition to the 24 hourly factors the socio-economic information collected at zip-code level from the US census data were used. This data was obtained from a SCDOT sponsored research project (Islam at al., 2017).

Input Feature- Categorical Features Preparation: Most AADT estimation models only used hourly volume (continuous features/variables) (Sharma et al. 1999 and Sharma et al. 2001). In this study, however, the models were developed with continuous and categorical features, specifically i) day of week and ii) month of year. Dummy variables were used for creating these categorical features. For developing the day of week variables, one feature was developed for each day for a total of 7 features for seven days in a week. For example, if a particular hourly volume set is for Monday, then the Monday features were assigned the value 1, and the features for the other days of the week were assigned 0. A similar method was used to develop the twelve month of the year categories.

Target Feature Features Preparation: The target feature used in this study is a factor of the actual AADT calculated at the ATR locations called AADT factor (equation 2).

$$\text{AADT factor} = \frac{AADT}{Sum\ of\ 24\ hourly\ volume\ of\ that\ day} \ldots\ldots\ldots (2)$$

For each ATR, the AADT is computed by calculating a simple average mean of all the available hourly volume for a year as mentioned in the Traffic Monitoring Guide (TMG, 2016).

3.2.3 Phase 3: Feature Selection

Feature selection was performed in order to reduce the use of irrelevant/insignificant features in developing either classification or prediction models, and to improve the model performance (Langley, 1994). In this study, two types of feature selection methods were performed. **Table 3-1** presents the feature selection methods applied for different types of data. The sequential feature selection method was used to select the best features from the 24 hourly volume. This method is a simple greedy search method which starts with an empty set of features. Eventually new features are added sequentially until the desired result from the criterion function is achieved.

Table 3-1 Feature Selection Methods

Features	Feature Selection Method
Continuous features: 24 hourly volume factors	Sequential Feature Selection
Other features: i) Income ii) Employment iii) Percent below poverty iv) Number of vehicles v) Number of housing units vi) Day of week and vii) Month of year viii) Number of lane	Cross Validation

The models developed for each of the functional class were run through the feature selection algorithm for selecting the best hourly volume factors resulting in the least residual sum of square errors. Once the best continuous features (hourly volume factors) were selected, the other features (census data and categorical) were combined to find the least error for predicting the target values/features using MATLAB.

3.2.4 Phase 4: AADT Estimation Model Development Using Artificial Intelligence

Once the continuous features were selected using the sequential feature selection method, and the other features (socio-economic variables and categorical features) were selected utilizing the cross validation method the models were developed using Artificial Neural Network (ANN) and Support Vector Regression (SVR). As mentioned earlier, separate models were developed for 5 functional classes of roadways of South Carolina

Each of the 5 functional class was then divided into different models based on the combination of different input features. **Table 3-2** presents the combination of the features in different candidate models for each functional class.

Table 3-2 List of Models and Input features for Different Functional Classes

Model	Input features
Model 1	Number of Lane, Day, Month, Income, Employment, Percent Below Poverty, Vehicles, Housing Unit, Hourly Volume Factors
Model 2	Day, Month, Hourly Volume Factors
Model 3	Vehicles, Housing Unit, Hourly Volume Factors
Model 4	Individual Day Model: Month, Hourly Volume Factors
Model 5	Individual Month Model: Day, Hourly Volume Factors

As mentioned earlier, the models were developed using two artificial intelligence paradigms, following sections discussed in detail how the models were developed using them.

Model Development Using Artificial Neural Network: Artificial Neural Network (ANN) is one of the most widely adapted alternatives to linear regression, logistic regression, time-series analysis, which are commonly used for developing predictive models (Tu, 1996). It has been used for successful pattern recognition, generalization and trend prediction (Sharma et al. 1999). In this study a multilayered, feed-forward, backpropagation neural network for supervised learning was used. The developed neural network model consists of three layers: the input layer, the hidden layer and an output layer. This ANN model is named as a feed-forward network as it feeds the output of one layer to another. A tan-sigmoid transfer function was used for calculating the output from each neuron. One of the remarkable characteristics of a back-propagation neural network is its ability to propagate the effects of error backward through the network after every training case (Leverington, 2009); thus this algorithm was chosen for estimating AADT. The training algorithm selected was the Levenberg-Marquardt, which is recommended for most of the prediction problems unless the data set is too noisy and small (Demuth et al., 1992). In this study, the author ran different ANN models with a different number of hidden neurons, with those neurons providing the least RMSE used for model development. The number of hidden neuron used in this study is varies based on models. **Figure 3-5** presents a sample neural network model, the calculation of the input and target features are detailed in section 4.2.1.

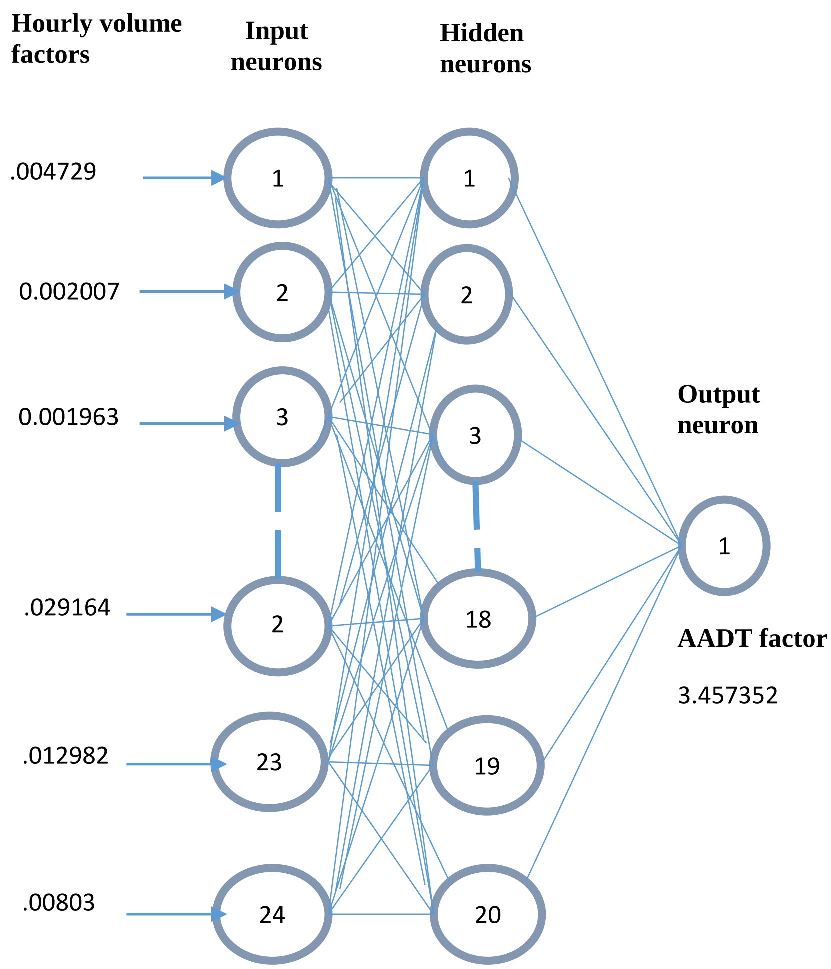

Figure 3-5 Sample Neural Network Model

Model Developed using Support Vector Regression: The SVM method has been successfully applied for classification and regression analysis via the construction of either one or more hyperplanes in a higher dimensional space. Developed as an extension of the nonlinear models of the generalized portrait algorithm, the SVM is based on the Vapnik-Chervonenkis (VC) and the statistical learning theories.

In order to perform the regression SVR executes two steps, first it performs nonlinear regression by mapping the training samples onto a high-dimensional, kernel-induced feature space. After that a liner regression is performed (Drucker, 1999). **Figure 3-6 presents an** overview of support vector regression.

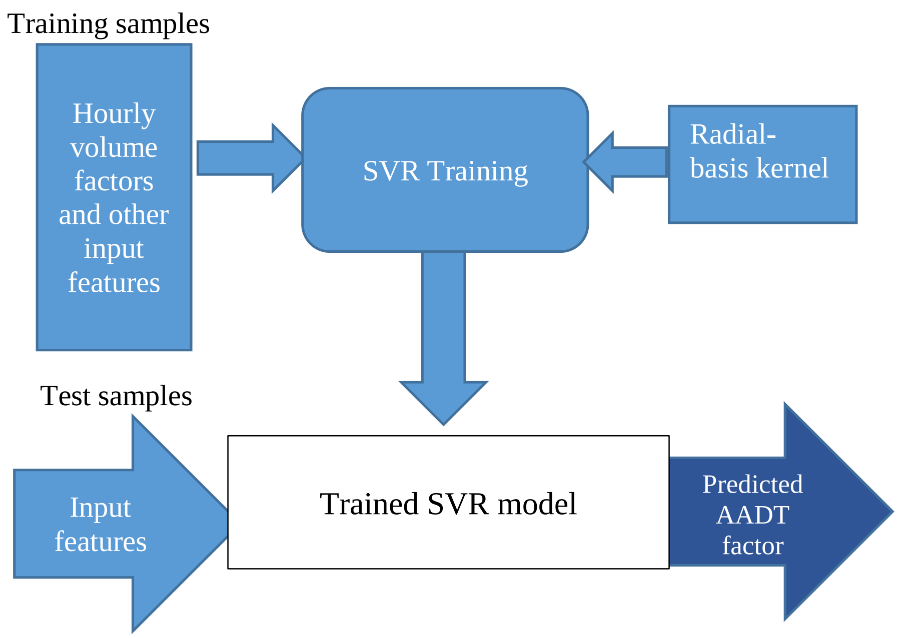

Figure 3-6 Overview of SVR model (Adopted from Bhavser et al. 2007)

Although the basic theories of SVR and SVM are very similar they have their differences too. In case of SVM there is a finite number of classifier but SVR has infinite number of target output within the training data. As a result, SVR tends to give any possible value in the output space from a group of input vectors.

In this study, MATLAB LIBSVM library tool in MATLAB (version 2013b) (Chang and Lin 2011) is used. The parameters used for SVR are C, γ and ε. C values varied for different combination of input features and for the models developed under different roadway functional classes. Different C and γ values were tested by increasing the value of n in exponential order (i.e., 2^n). The range of C is from 8 to 16 and the range for γ is -8 to 0 with a step of 2 increment. Once the C and γ values were determined using the grid search method and the ε value was found using cross validation. The value of the set of parameters varied from model to model with the change in training data.

3.3 Imputation of Missing Hourly Volume for ATRs Using Artificial Intelligence

This section outlines the method of developing missing hourly volume imputation models using Artificial Intelligence. The seven-phase method is presented in **Figure 3-6.** Each of the phases are described in detail in the following sections.

3.3.1 Phase 1 and 2: ATR selection and Data Collection

ATRs collect hourly volume 365 days a year. However, it was observed that there were a significant number of missing values in the collected data set at almost all ATRs. In this research, the author obtained hourly volume from 20 permanent count stations on the urban principal arterial- interstate, from 21 permanent count stations on the rural principal arterial- interstate and from 7 ATRs on urban principal arterial- other functional class of roadways for the year 2014. The hourly volume from different permanent count stations were collected from SCDOT. The data in this database did not contain any type of imputation or manipulation of hourly volume. Similar to data used for developing the AADT estimation models, two types of input features were used for missing hourly volume imputation models:

a) Hourly volume available before the missing hours data

b) Categorical data: day of week, month of year and direction of traffic.

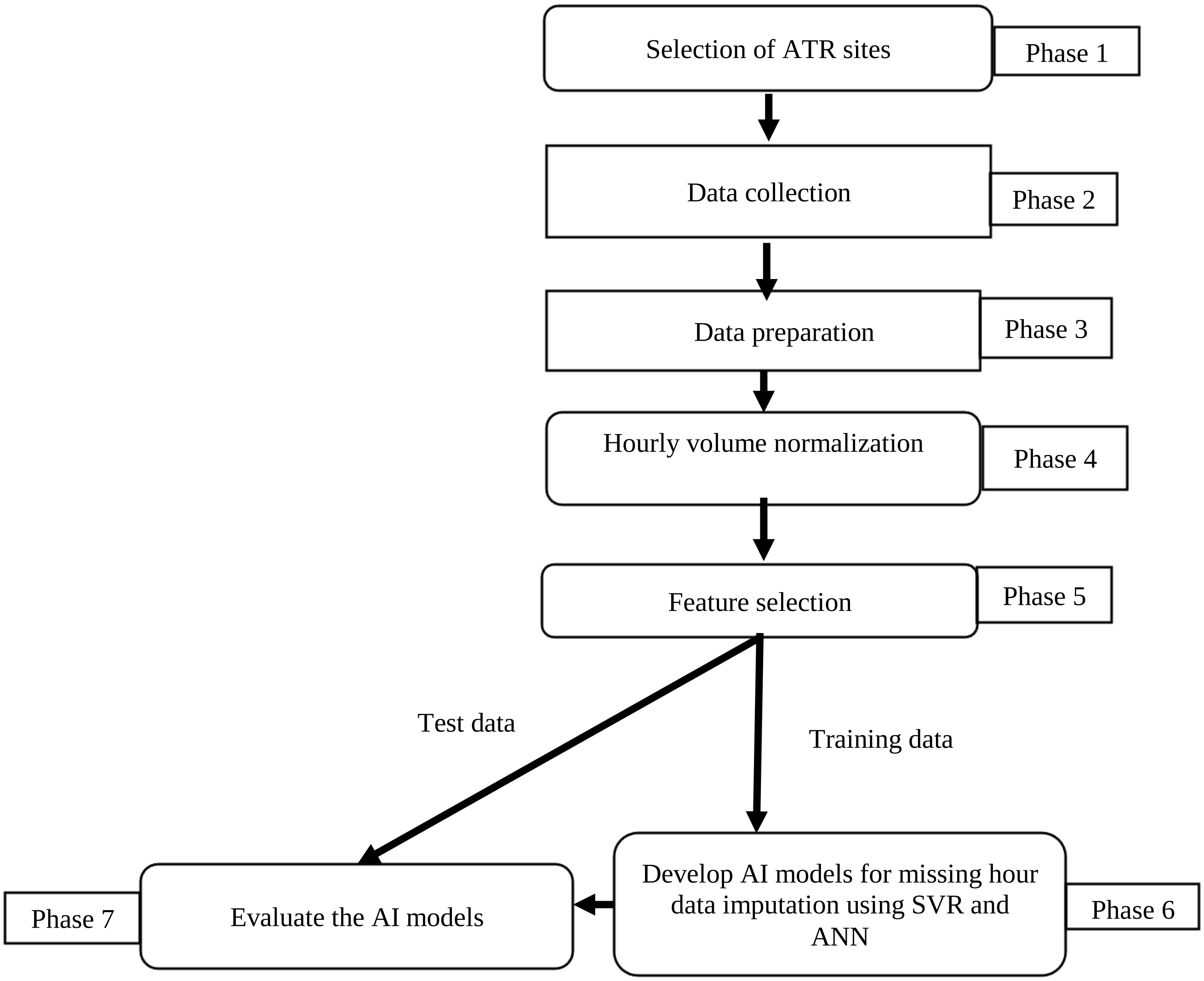

Figure 3-6 Method for missing hourly volume imputation

3.3.2 Phase 3: Data Preparation

Data preparation is one of the most important steps for developing models. For developing these models, the following features were used

a) input features: hourly volume before the missing hours (12AM-12AM data before the assumed missing hour data) and categorical features (Day of week, Month of Year and Direction of Traffic)

b) target feature: Hourly volume that was missing from the permanent count stations. While developing the models it was assumed continuous 8 hours of data were missing, hence the target feature was the hourly volume from the hour 12AM-8AM. Section 4.3.1 presents how the input and target features are selected.

3.3.3 Phase 4: Hourly Volume Normalization

Once the hourly volume was prepared in the previous phase, it was necessary to normalize data (both input hourly volume features and the target features) for the models developed under different functional classes. The data were normalized using the following formula:

$$\text{Normalized hourly volume} = \text{Absolute } \left(\frac{X - X_{mean}}{std\ (x)}\right)$$

X= hourly volume for a particular hour

X_{mean}= mean of the hourly volume for a particular hour for a year

$std\ (x)$ = Standard deviation of the hourly volume for a particular hour for a year

3.3.4 Phase 5: Feature Selection

Feature selection methods were applied to select the significant features for the missing hourly volume imputation models. The hourly volume features were selected using the sequential feature selection method. Once the best hourly volume features were selected, different combinations of the categorical features were combined to find the combination resulting in the least RMSE values.

3.3.5 Phase 6: Model Development Using Machine Learning Techniques

As discussed in the previous section, a combination of the hourly volume and the categorical data were prepared for different functional classes. The models were developed for the following roadway functional classes

a. Urban Principal Arterial- Interstate and Expressways

b. Rural Principal Arterial- Interstate

c. Urban Principal Arterial – Other

In this study, for each of the three functional class of roadways, following 4 models were developed to determine the model with least RMSE error.

Table 3-3 List of Models and Input features for Different Functional Classes

Model	**Input features**
Model 1	Day, Month, Hourly Volume Available Before the Missing Hours
Model 2	Day, Month, Hourly Volume Available Before the Missing Hours
Model 3	Individual Day Model: Month, Hourly Volume Available Before the Missing Hours
Model 4	Individual Month Model: Day, Hourly Volume Available Before the Missing Hours

The next step prior to model development is separating the data into train and test cases. Similar to the models developed for AADT estimation, 2/3 of the data from the entire data set were used for training and development of the learning algorithm and 1/3 of the data were used for testing the developed algorithms.

Model Development Using Artificial Neural Network: A multilayer feed forward neural network with back propagation learning was used for developing the missing hourly volume imputation models. The developed neural network consists of three layers: a) an input layer; b) a hidden layer; and c) an output layer). As this is a backpropagation algorithm, it has the ability to propagate the effects of error backward through the network after every training case, and this characteristic of the network to adjust error is one of the motivating factors for choosing this particular architecture of ANN for missing data imputation. The training algorithm used is Levenberg-Marquardt. In this study, trial and error method was performed to find the number of neurons that produce the minimum RMSE. The neural network model was implemented in MATLAB using the library function NNtool (Demuth, 1992).

Model Development Using Support Vector Regression: In this study, a support Vector regression algorithm with radial basis kernel function was chosen from the MATLAB LIBSVM library tool in MATLAB (version 2013b) (Chang and Lin 2011). The parameters used for SVR are C, γ and ε. C values varied for different combination of input features and for the models developed under different functional classes. Different C and γ values were tested by increasing them in exponential order. i.e. 2^n, in the range of 8 to 16 for C and -12 to -4 for γ with a step of 2. Once the C and γ values were determined using the grid search method the ε value was found using cross validation. The value of the set of parameters varied from model to model with the change in training data.

CHAPTER FOUR

ANALYSIS AND RESULTS

4.1 Overview

This chapter presents the results and analysis of the following two primary sections:

1. **Estimation of Annual Average Daily Traffic (AADT)**
   a. Evaluation of the estimated AADT using the artificial intelligence (AI) models developed with two machine leaning techniques (SVR and ANN) (section 4.2.1)
   b. Comparison of the estimated AADT using machine leaning techniques to Traditional Factor method used by SCDOT (section 4.2.2)
   c. Comparison of the estimated AADT using machine leaning techniques to an Ordinary Least Square Regression based method (section 4.2.3)
2. **Imputation of Missing Hourly Volume from the ATR Stations**
   a. Evaluation of the imputed missing hourly data using the models developed with two machine leaning techniques (SVR and ANN) (4.3.1)
   b. Comparison of the imputed hourly volume using machine leaning techniques to the historical average method used by SCDOT (4.3.2)

4.2 Evaluation of AI Models for Estimating Annual Average Daily Traffic

This section presents the performance evaluation of the Artificial Intelligence (AI) models developed using two machine learning techniques. After that, the AADT estimated by the best AI models are compared to the AADT estimated by the traditional factor method used by SCDOT. In addition, a comparison is conducted between the AADT estimated by the AI models and a regression based method.

The performance of models is decided based on the Root Mean Square (RMSE) and Mean Average Percentage Error (MAPE) values. The formulas used for calculating RMSE and MAPE are given below

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\left(\frac{\sum_{i=1}^{n}(Yi-yi)^2}{n}\right)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$$\text{Mean Average Percentage Error (MAPE)} = \frac{1}{n}\sum\left(\frac{|Yi-yi|}{Y}\right) * 100 \ldots\ldots\ldots(2)$$

For i_{th} day,

Y_i = Actual AADT

y_i = Predicted AADT

n = Number of observations

4.2.1 Evaluation of Estimated AADT using Machine Learning Techniques

In this section the evaluation of AI models developed for the 5 roadway functional classes (as discussed in section 3.2.4) is presented. Prior to present the results from the models, the steps performed for developing the models are discussed.

4.2.1.1 Input and target feature calculation for ANN and SVR models

Section 3.2.2 presents the formula for calculating the input features and target feature. In this section, a sample calculation of these features for one of the ATRs from principal arterial is presented (Please see **Table 4-1**).

Input features calculation

For an ATR in the Principal Arterial (ATR ID - 6) (Date: 01/03/2011)

AADT = 77,500 and

Sum of 24-hour volume from the day (01/03/2011) = 22,416

Volume for 1AM- 2 AM (Both direction) = 106 veh

Volume for 2AM- 3 AM (Both direction) = 45 veh

Volume for 3AM- 4 AM (Both direction) = 44 veh

So the hourly volume factors (Input feature) are:

Hourly volume factor for 12 AM- 1 AM (Both direction) = 106/ 22416 = 0.004729 (Column 26)

Hourly volume factor for 1 AM- 2 AM (Both direction) = 291 / 22416= 0.002007 (Column 27)

Hourly volume factor for 2 AM- 3 AM (Both direction) = 257 / 22416 = 0.001963 (Column 28)

Column 9 to Column 25 (All columns are not shown in the figure) in Table 4-1 represents the categorical features. The date 1/3/2011 is a Monday, so the column for Monday (column 9) is assigned 1 and categorical features related to other days are assigned zero. Similarly, as the data is for January, column 15 for January is assigned 1 and the rest of the columns for the other 11 months are assigned 0 (Table 4-1 only shows the month January-March and Hourly volume from Hour 1-Hour 3 and Hour 23-Hour 24).

Apart from the hourly volume features and the categorical features, there are also socio-economic features (Not shown in this table) listed in section 3.2.1. Also Appendix B contains the list of socio-economic features used in this study. The number of lane for this ATR is 4

Target features calculation

The target feature, AADT factor for Monday is calculated using the following formula:

AADT factor = 77500 / 22416 = 3.457352

Table 4-1 presents the sample input and target features used for developing different AADT estimation models listed in **Table 3-2**

Table 4-1 Input and Target Features of AADT Estimation Models

		Input Features													Target Feature
Column 1	Column 2	Column 3	Column 9	Column 10	Column 12	Column 13	Column 14	Column 15	Column 16	Column 17	Column 26	Column 27	Column 28	Column 45	Column 46	Column 47
		Input Features													Target Feature
			Categorical Features								Hourly volume features				
ATR ID	Date	No of Lane	Monday	Tuesday	Friday	Saturday	Sunday	Jan	Feb	Mar	Hour1 factor	Hour2 factor	Hour3 factor	Hour23 factor	Hour24 factor	AADT factor
6	1/3/2011	4	1	0	0	0	0	1	0	0	0.0047	0.002	0.002	0.013	0.008	3.4574
6	1/10/2011	4	1	0	0	0	0	1	0	0	0.0069	0.0044	0.0038	0.017	0.0102	7.3321
6	1/17/2011	4	1	0	0	0	0	1	0	0	0.0059	0.0031	0.0023	0.0115	0.0088	3.7875
6	1/24/2011	4	1	0	0	0	0	1	0	0	0.0031	0.0013	0.0017	0.0101	0.0065	3.0311
6	1/31/2011	4	1	0	0	0	0	1	0	0	0.0039	0.0023	0.0018	0.0106	0.006	3.0819
6	2/7/2011	4	1	0	0	0	0	0	1	0	0.0041	0.002	0.0014	0.0095	0.0062	3.2404
6	2/14/2011	4	1	0	0	0	0	0	1	0	0.0029	0.0019	0.001	0.0167	0.0079	2.6641
6	2/21/2011	4	1	0	0	0	0	0	1	0	0.0044	0.0028	0.0009	0.0113	0.007	3.0135
6	2/28/2011	4	1	0	0	0	0	0	1	0	0.0035	0.002	0.0014	0.0116	0.0064	2.7700
6	3/7/2011	4	1	0	0	0	0	0	0	1	0.0035	0.0016	0.0015	0.0111	0.0074	2.7778
6	3/21/2011	4	1	0	0	0	0	0	0	1	0.004	0.0019	0.0012	0.0135	0.0074	2.6652

4.2.1.2 Parameter adjustment for SVR method

Accurate estimation of the SVR parameters are the key for correct prediction of AADT. It is mentioned in the method section that both cost coefficient (C) and the kernel parameter (γ) are estimated using the grid search method. The optimal value of C and γ parameters are chosen based on the highest cross-validation accuracy using the training data. The epsilon (ε) values varied between 0.00001to 0.000075, which was determined based on cross validation method. Using the optimal values of these SVR parameters, trained SVR model files are generated in MATLAB to estimate AADT for the test cases. **Table 4-2** shows the optimal values of SVR parameters with least RMSE for different roadway functional classes.

4.2.1.3 Number of hidden neuron determination for ANN

While developing the neural network models for estimating AADT, the number of hidden neurons played an important role for prediction. It is mentioned in the method section that the number of hidden neuron is determined based on cross validation. The number of hidden neurons of the ANN models for estimating AADT varied between 5-20.

Table 4-2 SVR Parameter Values with least RMSE

SVR Parameters	**Urban Principal Arterial-Interstate (Model 5)**	**Rural Principal Arterial-Interstate (Model 5)**	**Urban Principal Arterial-Other (Model 3)**	**Rural Principal Arterial-Other (Model 2)**	**All Functional Class (Model 5)**
C	2000	2000	2000	2000	2000
ϒ	.5	.5	.5	1	.5

4.2.1.4 Selected Features for Developing AI Models

The method of feature selection is described in greater depth in section 3.2.3. **Table 4-3** presents the number of hourly volume features selected out of the 24 available hourly volume features using the sequential feature selection method for different roadway functional classes. This table also presents the total number of features of the models with least RMSE for different roadway functional classes. The total number of features include the categorical and socio-economic and hourly volume features depending on the model.

Table 4-3 Features selected using Sequential Feature Selection Method and the total number of features of the models with least RMSE

Feature Type	Urban Principal Arterial-Interstate (Model 5)	Rural Principal Arterial-Interstate (Model 5)	Urban Principal Arterial-Other (Model 3)	Rural Principal Arterial- Other (Model 2)	All Functional Class (Model 5)
Selected hourly volume features	13	11	21	14	19
Total Features	20	18	42	33	26

4.2.1.5 Model Evaluation: Urban Principal Arterial- Interstate

SCDOT has most of its permanent count stations in the higher functional class of roadways, and Urban Principal Arterial-Interstate is one of them. The models are developed for this functional class group using 20 ATR stations. In order to keep the training data set separate from the testing data set, 13 ATRs (two third of the data set) were used for training and the remaining 7 ATRs (one third of the data set) were used for the test to predict AADT based on the trained model. Under this functional class group different combination of features were tested to find a combination that can estimate AADT with least errors. The errors are calculated by using the actual AADT factors of ATRs with the estimated AADT factors from the AI models. **Table 4-4** presents the RMSE of Urban Principal Arterial-Interstate group model for different combination of input features. **Figure 4-1** illustrates a graphical representation of the errors for five models. Please see appendices for the detailed RMSE calculation.

Table 4-4 RMSE of Urban Principal Arterial – Interstate Models

Models	Input Features	RMSE (SVR)	RMSE (ANN)
Model 1	Number of Lane, Day, Month, Income, Employment, Percent Below Poverty, Vehicles, Housing Unit, Hourly Volume Factors	0.3927	0.4113
Model 2	Day, Month, Hourly Volume Factors	0.3824	0.4914
Model 3	Vehicles, Housing Unit, Hourly Volume Factors	0.3906	0.3942
Model 4 (Monday)	**Individual Day Model:** Month, Hourly Volume Factors	0.3208	0.9891
Model 5 (January)	**Individual Month Model:** Day, Hourly Volume Factors	0.3168	0.3372

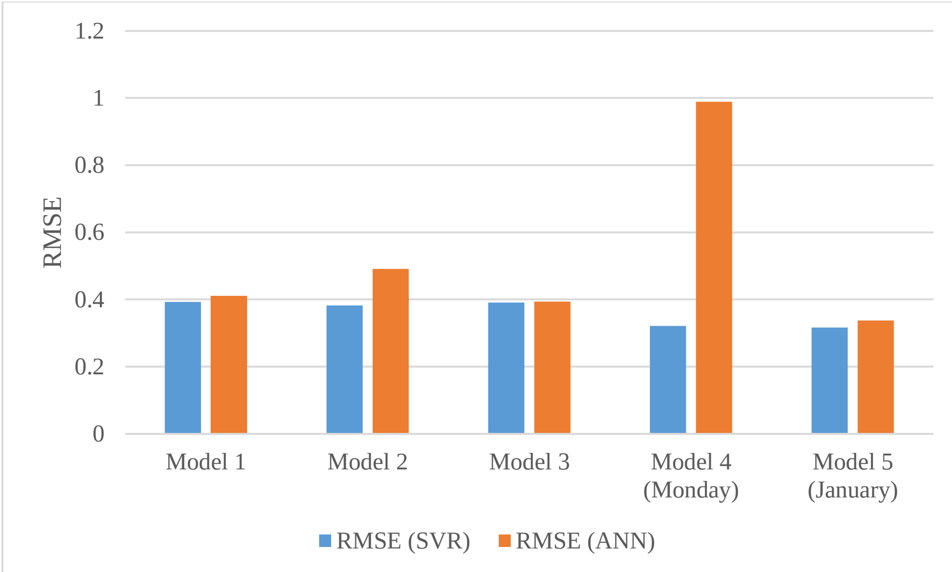

Figure 4-1 RMSE of Urban Principal Arterial – Interstate Models

In order to test if the predicted AADT factors are significantly different from the actual AADT factors, Z tests were conducted. The results from the tests indicated that, SVR – model 3, SVR – model 5 and ANN – model 5 predicted AADT factors that are not significantly different from the actual AADT factors at 95% level of confidence. Each of the 5 models consists of different combinations of input features such as the hourly volume factors, socio-economic variables, and other categorical features (day of week, month). Since the SVR model can guarantee global minima for a given set of training data, it is expected to perform better for prediction (Wu et al. 2004). In terms of the RMSE, it can be said that SVR has least RMSE than ANN for each of the models. It is also evident that the SVR performance increased (with decrease in RMSE) in individual day and month models (model 4 and model 5). The reason for this better performance is the similarity in traffic volume in these models which eases the prediction of AADT. A comparison of the model errors shows that the addition of socio-economic features with hourly volume features (mode l1 and model 3) did not improve the model performance.

4.2.1.6 Model Evaluation: Rural Principal Arterial- Interstate

The AADT estimation models for Rural Principal Arterial- Interstate group were developed using 24 ATRs. The models consist of 11 hourly volume factors which were selected using feature selection method out of the 24 hourly volume factors, and other socio-economic and categorical features. The error estimation for five models are presented in **Table 4-5,** and are illustrated in **Figure 4-2**.

Table 4-5 RMSE of Rural Principal Arterial – Interstate Model

Models	Input Features	RMSE (SVR)	RMSE (ANN)
Model 1	Number of Lane, Day, Month, Income, Employment, Percent Below Poverty, Vehicles, Housing Unit, Hourly Volume Factors	0.3553	0.3704
Model 2	Day, Month, Hourly Volume Factors	0.2085	0.2224
Model 3	Vehicles, Housing Unit, Hourly Volume Factors	0.3529	0.3549
Model 4 (Monday)	**Individual Day Model:** Month, Hourly Volume Factors	0.2319	0.2655
Model 5 (January)	**Individual Month Model:** Day, Hourly Volume Factors	0.1992	0.2939

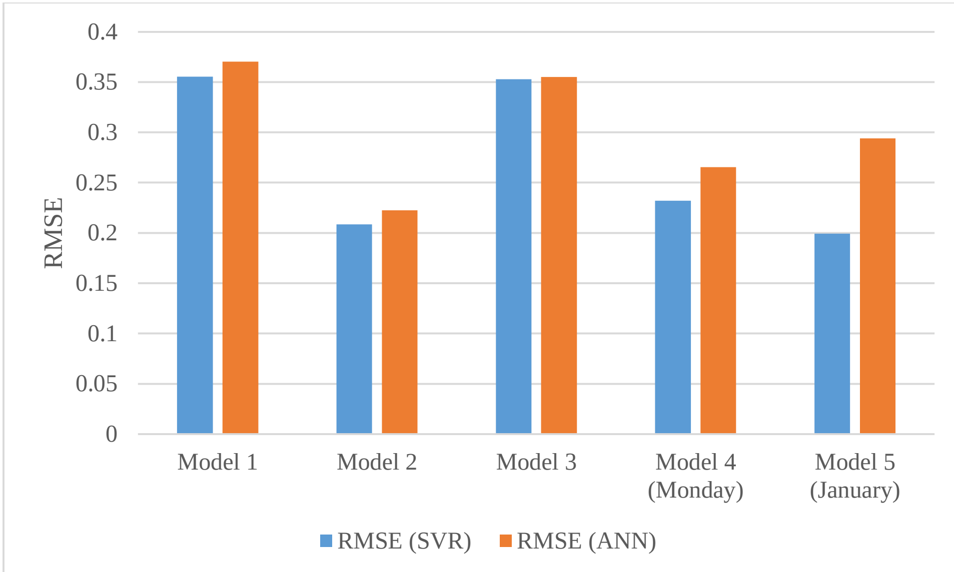

Figure 4-2 RMSE of Rural Principal Arterial – Interstate Models

Analyzing the results from Table 4-5, it is evident that SVR continued to perform better in each of the models for predicting the AADT factors. In order to test if the predicted AADT factors are significantly different from the actual AADT factors, Z tests were conducted. The results from the tests indicated that, SVR – model 1, model 4 and model 5; ANN – model 3, model 4 and model 5 predicted AADT factors that are not significantly different from the actual AADT factors at 95% level of confidence. Among these models SVR – model 5 resulted the least RMSE value.

4.2.1.7 Model Evaluation: Urban Principal Arterial- Other

Following the functional class division by SCDOT, this model group for AADT estimation is developed utilizing 8 permanent count stations. 6 ATRs were used for training the model and 2 were used for testing it. **Table 4-6** presents the RMSE values of each of the models developed for this functional class group and **Figure 4-3** shows the graphical representation of the RMSE values.

Table 4-6 RMSE of Urban Principal Arterial – Other Model

Models	Input Features	RMSE (SVR)	RMSE (ANN)
Model 1	Number of Lane, Day, Month, Income, Employment, Percent Below Poverty, Vehicles, Housing Unit, Hourly Volume Factors	0.6286	0.630
Model 2	Day, Month, Hourly Volume Factors	0.2779	0.3138
Model 3	Vehicles, Housing Unit, Day, Month Hourly Volume Factors	0.2116	0.4858
Model 4 (Monday)	**Individual Day Model:** Month, Hourly Volume Factors	0.4411	0.7131
Model 5 (January)	**Individual Month Model:** Day, Hourly Volume Factors	0.4761	1.0806

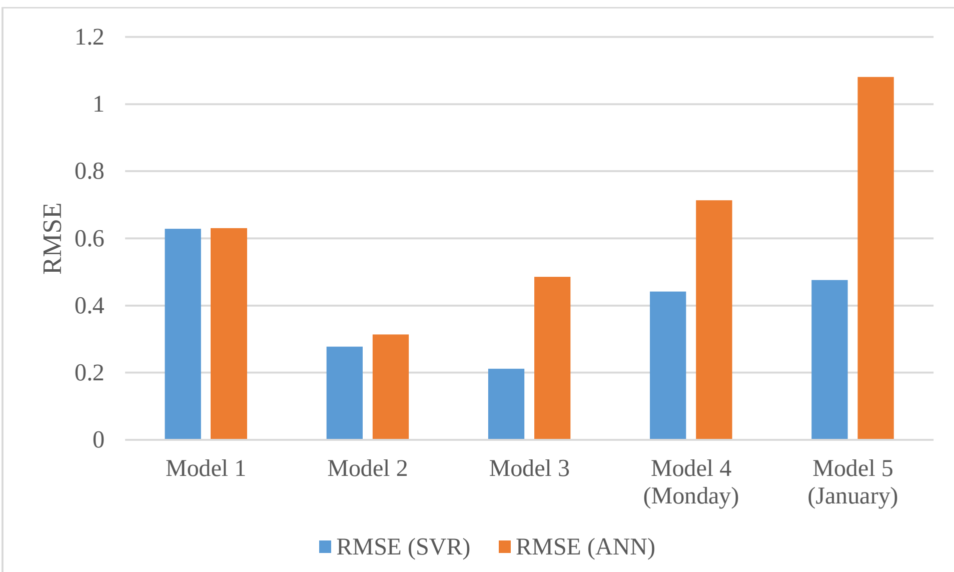

Figure 4-3 RMSE of Urban Principal Arterial – Other Models

model 1 generates the highest RMSE values for SVR which depicts that adding the socio-economic variables did not add any values to predicting AADT. In order to test if the predicted AADT factors are significantly different from the actual AADT factors, Z tests were conducted. The results from the tests indicated that, SVR – model 4 and model 5; ANN – model 3, model 4 and model 5 predicted AADT factors that are not significantly different from the actual AADT factors at 95% level of confidence. The errors of the SVR method depends on the accurate estimation of the SVM parameters. Both the cost coefficient (C) and the kernel parameter γ are estimated using the grid search method. The optimal value of C and γ parameters are chosen based on the highest cross-validation accuracy.

4.2.1.8 Model Evaluation: Rural Principal Arterial- Other

This functional class group models are developed using 20 permanent count stations. 13 ATRs were used for training and rest were used for testing the trained models. **Table 4-7** presents the RMSE value of each of the models developed for this functional class and Figure 4-4 presents the graphical representation of the RMSE for the rural principal arterial-other. The results from the Z test revealed that, SVR - model 4 and ANN – model 4 predicted AADT factors that are not significantly different from the actual AADT factors at 95% level of confidence.

Table 4-7 RMSE of Rural Principal Arterial – Other Model

Models	Input Features	RMSE (SVR)	RMSE (ANN)
Model 1	Number of Lane, Day, Month, Income, Employment, Percent Below Poverty, Vehicles, Housing Unit, Hourly Volume Factors	0.3974	0.4399
Model 2	Day, Month, Hourly Volume Factors	0.2420	0.3161
Model 3	Vehicles, Housing Unit, Hourly Volume Factors	0.3973	0.3478
Model 4 (Monday)	**Individual Day Model:** Month, Hourly Volume Factors	0.2786	0.3278
Model 5 (January)	**Individual Month Model:** Day, Hourly Volume Factors	0.5291	0.6369

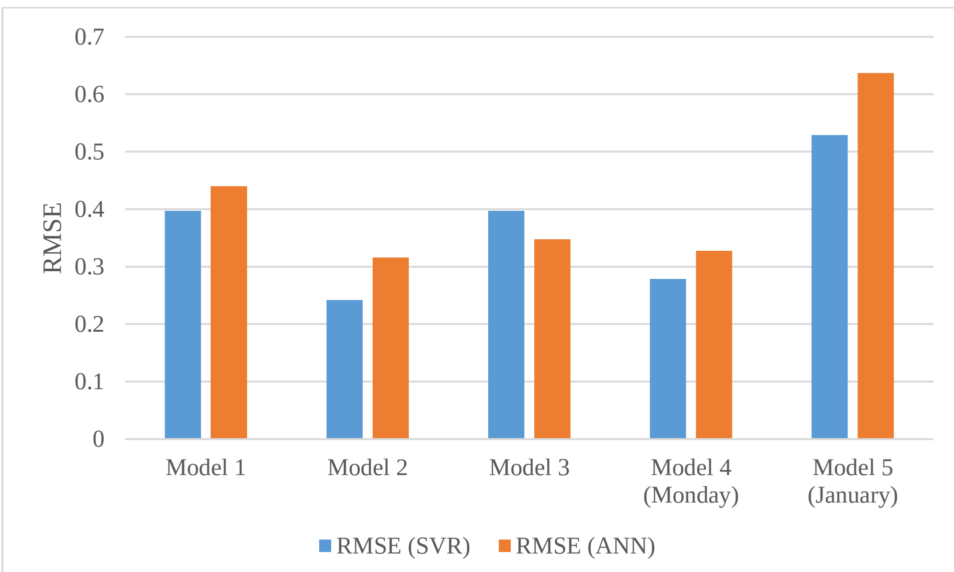

Figure 4-4 RMSE of Rural Principal Arterial – Other Models

4.2.1.9 Model Evaluation: General Model

This general model includes all ATRs. The training features used for developing models are the hourly volume factors, month of the year, and day of week. The RMSE of different models estimated using 117 ATRs of South Carolina is presented in **Table 4-8.** Unlike the functional class specific models, ANN predicted the AADT factors better than SVR for this general model. Results from the Z test revealed that, SVR – model 5 and ANN – model 5 predicted AADT factors that are not significantly different from the actual AADT factors at 95% level of confidence. This model has the potential to predict AADT factors irrespective of the functional class of ATRs. **Figure 4-5** shows the graphical representation of the RMSE values of different models.

Table 4-8 RMSE of General Model

Models	Input Features	RMSE (SVR)	RMSE (ANN)
Model 2	Day, Month, Hourly Volume Factors	0.3461	0.3551
Model 4 (Monday)	**Individual Day Model:** Month, Hourly Volume Factors	0.4586	0.3232
Model 5 (January)	**Individual Month Model:** Day, Hourly Volume Factors	0.3342	0.3133

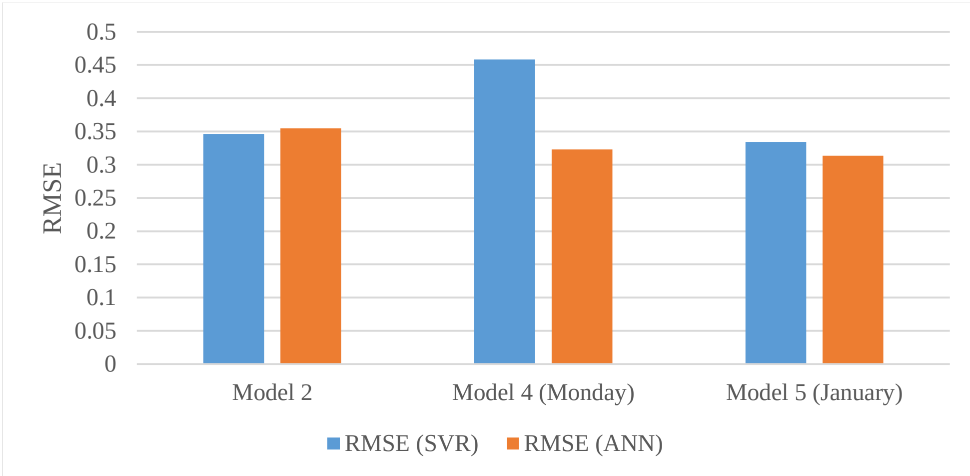

Figure 4-5 RMSE of All ATR Functional Class Models

4.2.2 Comparison Between Support Vector Regression and Traditional Factor Method Performance

One of the objectives of this study was to find the efficacy of the models developed using the machine learning techniques over the traditional factor method used by SCDOT. In traditional factor method for estimating AADT, SCDOT uses two types of factors

1. Seasonal or monthly factors
2. Axle correlation factor

These factors are calculated for each of the roadway functional class. Then the short term counts conducted in these functional classes are multiplied with these functional class specific factors to estimate AADT. This section presents the comparison between the AADT estimated by SVR with the traditional factor method used by SCDOT. Between the two AI paradigms, SVR is chosen for comparison because SVR predicted AADT better than ANN. For comparing the AADTs predicted by SVR and factor method, different days were chosen which were assumed as different short term counts for different times of the year. For predicting AADT factors using SVR, hourly volume factors and other factors were used for the selected day. The predicted AADT factor was multiplied with sum of 24 hourly volumes to calculate the AADT. To predict AADT using factor method, the sum of 24-hour volume for the selected day was multiplied with the monthly factor and seasonal factor. In this section the AADT values are compared for urban and rural principal arterial – other roadway functional classes. **Table 4-9** presents the actual AADT and predicted AADT by the two methods. The R^2 values for the two models are presented in **Figure 4-6**. From the figure it can be seen that SVR was producing models with higher R^2 (.8452) compared to the traditional factor method (R^2=.8094). Also the MAPE value was lower for SVR (16.32%) than the factor method (21.22%).

Table 4-9 Comparison of AADT estimated by SVR to Traditional Factor Method

Actual AADT	Estimated AADT(SVR)	Estimated AADT (SCDOT method)	MAPE(%) of SVR	MAPE(%) of factor method
16400	16304	15576	0.586	5.024
16400	12782	2701	22.063	83.530
16400	15559	16520	5.129	0.732
16400	15091	17794	7.983	8.498
16400	15487	16412	5.566	0.075
2000	1260	1084	36.994	45.791
2000	2145	2146	7.243	7.319
41200	26355	24072	36.032	41.574
41200	41196	34935	0.011	15.206
41200	53750	46076	30.460	11.836
41200	54531	46930	32.357	13.908
		Total	16.766	21.227

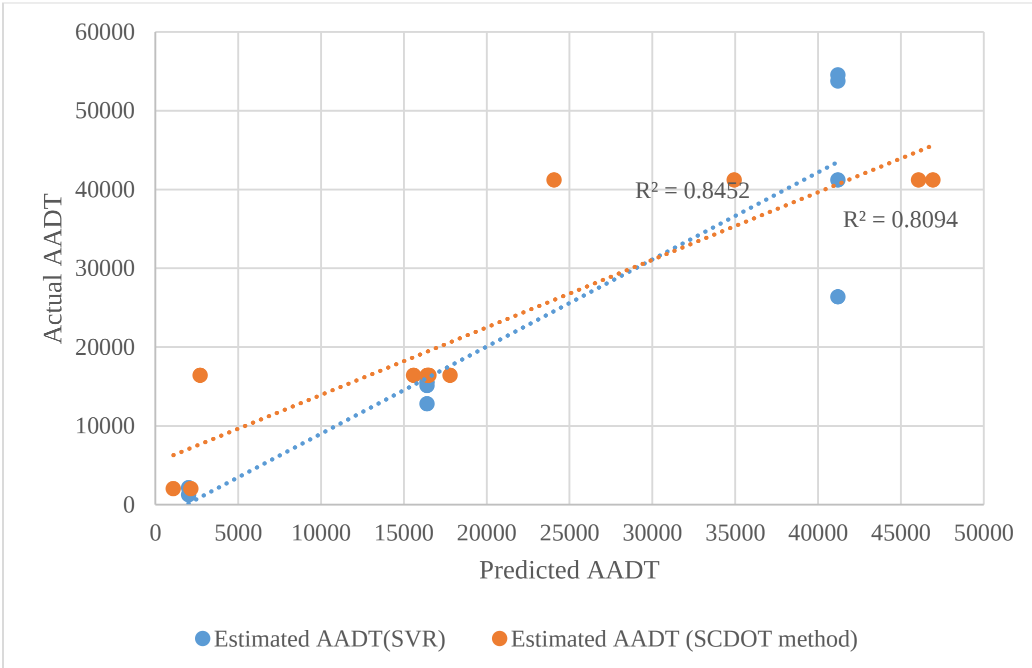

Figure 4-6: R^2 for the SVR and Factor Method

4.2.3 Comparison Between Support Vector Regression and Ordinary Least Square Regression Method

This section presents the comparison between AADT estimated using SVR and an Ordinary Least Square Regression Method. This regression model consists of ATRs from both Principal Arterials (Interstates) and Minor Arterials for urban and rural roadways. The regression model was developed for a research project sponsored by SCDOT. As the previous models presented in this study were functional class specific, for the comparison purpose, SVR models were developed combining the principal and minor arterials. Both regression and the AI models were developed using 47 permanent count stations. Among the five AI models developed, model 2 was the model with the least RMSE. A paired t-test (at a 95% confidence level) of the differences between the actual and SVR output indicated no statistical difference between the Actual and SVR predicted AADTs. There was, however, significance difference between actual AADT and AADT estimation using the regression method. In addition, R^2 values and MAPE (%) were next calculated to compare the performance of the both models, and presented in **Figure 4-7.** In terms of MAPE, SVR model performed better compared to the regression model (i.e., lower MAPE (6.817) value of the SVR than the MAPE value (45.267) of regression model.

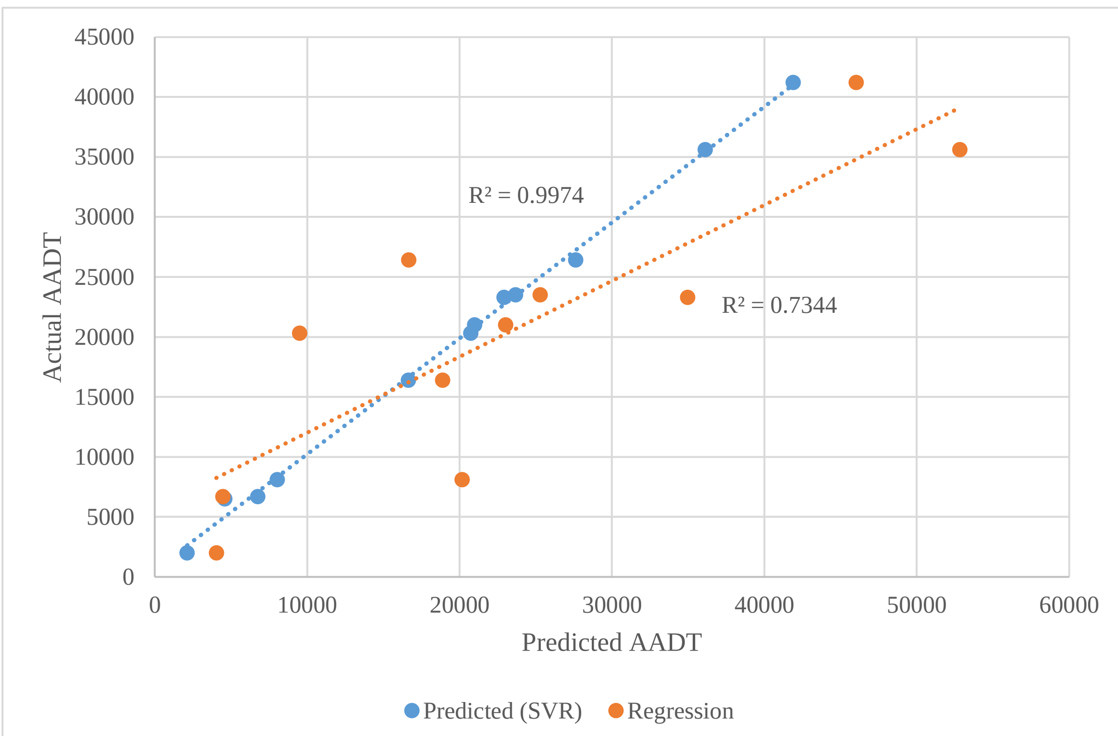

Figure 4-7: R^2 for the SVR and Regression Model

4.3 Evaluation of Models for Imputing Missing Hourly Volume

In order to impute the missing hourly volume from the permanent count stations, models were developed for 3 different functional classes of roadways using machine learning techniques (i.e., ANN and SVR). The results were compared with the traditional historical average method used by the SCDOT for imputing missing hourly volume. Similar to the AADT estimation method, the evaluation criteria were the RMSE and MAPE (%) of the developed models. Prior to present the results from the models, the steps performed for obtaining the results are discussed.

4.3.1 Evaluation of missing hourly volume imputation using Machine Learning Techniques

The preparation of the data for developing models was discussed in details in Section 3.3.2. In this section an illustrative example of how input and target features are chosen is provided.

Table 4-10 Input (hourly volume only) and Target Feature Determination

		1AM-12AM (24 hours data)											
ATR ID	**Date**	**1:00 AM**	**2:00 AM**	**3:00 AM**	**4:00 AM**	**5:00 AM**	**6:00 AM**	**7:00 AM**	**8:00 AM**	**9:00 AM**	**10:00 PM**	**11:00 PM**	**12:00 AM**
23	1/1/2014	420	330	258	154	104	164	239	305	393	755	465	408
23	1/2/2014	233	155	145	158	251	444	765	1116	1069	867	605	446

While developing the missing hourly data imputation models, for an ATR No. 23 from the Urban Principal Arterial it was assumed that on 1/2/2014, the hourly volume from 1AM to 8AM were missing (Highlighted with green in Table 4-10). However, 24 hourly volumes for pervious day (1/1/2014) were available. Now to impute missing data for these 8 hours for the day 1/2/2014, 8 different models were prepared for each hour.

To impute the missing hourly volume for the hour 12 AM to 1 AM on 1/2/2014, the input features are the hourly volumes from 12 AM to 12 AM (420, 330,258,408) on 1/1/2014, where the target volume/feature is 233 veh.

Similarly, for, imputing the missing hourly volume for the hour 1 AM to 2 AM on 1/2/2014, the input features remain the same: the volumes from 12 AM to 12 AM (420, 330,258…….408) on 1/1/2014 and the target feature is 155 veh.

This procedure continues for the rest of the assumed missing hours.

The categorical feature creation is similar to the procedure described in section 4.2.1 for the AADT estimation models.

4.3.1.1 Parameter adjustment for SVR method

Accurate estimation of the SVR parameters are the key for correct prediction of missing hourly volume. As discussed in the method section, both cost coefficient (C) and the kernel parameter γ are estimated using the grid search method. The optimal value of C and γ parameters are chosen based on the highest cross-validation accuracy of the training data. After the optimization, cross validation was applied to the parameters to get higher accuracy. Using the optimal values of these SVR parameters, trained SVR model files are generated in MATLAB to estimate missing hourly volume for the test cases. **Table 4-11** shows the optimal values of SVR parameters for the best model developed for different roadway functional classes. The epsilon (ε) values varied between .0001 to .0005. The value was determined based on cross validation.

Table 4-11 SVR Parameter Values

SVR Parameters	Urban Principal Arterial- Interstate	Rural Principal Arterial- Interstate	Urban Principal Arterial- Other
C	20000	20000	20000
ϒ	.0005	.0005	.0005

4.3.1.2Number of hidden neuron determination for ANN Method

While developing imputation model using the neural networks, the number of hidden neurons played an important role for prediction. It is mentioned in the method that the number of hidden neuron is determined based on cross validation. For the models developed for missing hourly volume imputation, the number of hidden neuron varied between 5-35.

4.3.1.3 Selected Features for Developing AI Models

It is mentioned earlier in section 3.3.4 that two types of feature selection method had been applied to the features. **Table 4-12** presents the number of hourly volume features selected out of the 24 available hourly volume features using the sequential feature selection method for different roadway functional classes. Also the table presents the number of total features of the models that generated the least RMSE values. The number of total feature consists of hourly volume features and the categorical features.

Table 4-12 Features selected using Sequential Feature Selection Method and the total number of features of the models with least RMSE

Types of Feature	Urban Principal Arterial- Interstate	Rural Principal Arterial- Interstate	Urban Principal Arterial- Other
Selected hourly volume features	13	20	16
Total Features	29	31	35

4.3.1.4 AI Model Evaluation: Urban Principal Arterial- Interstate

In order to impute missing hourly volume from the permanent count stations of urban principal arterials, models were developed utilizing 21 permanent count stations. 2/3 of the data from the entire data sets were used for training and the rest were used for testing. It was assumed that for the ATRs, hourly volume was missing for up to 8 hours. The models were developed for 8 hours because data obtained from SCDOT revealed that the data base had data missing from 1 hour to 8 hours most of the time. However, SCDOT does not impute missing hourly volume if data for 12 consecutive hours are missing for one day.

The root mean square error values generated from each of the models for different hours for imputing missing hourly volume for urban principal arterial-interstate are presented in **Table 4-13.** Figure 4-8 and 4-9 presents the graphical presentation of the errors for SVR and ANN and Figure 4-10 shows the graphical representation of the average RMSE of ANN and SVR.

Table 4-13: RMSE of Urban Principal Arterial – Interstate Model

		RMSE (SVR)								
		Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7	Hour 8	Average
Model 1	Day, Month, Hourly Volume Available Before the Missing Hours	0.382	0.632	0.909	1.009	0.779	0.442	0.309	0.290	0.594
Model 2	Day, Month, Hourly Volume Available Before the Missing Hours	0.359	0.639	0.913	1.011	0.775	0.445	0.310	0.290	0.593
Model 3(Monday)	Month, Hourly Volume Available Before the Missing Hours	0.195	0.344	0.462	0.380	0.350	0.378	0.360	0.409	0.360
Model 4 (January)	Day, Hourly Volume Available Before the Missing Hours	0.964	1.599	2.184	2.158	1.408	0.695	0.367	0.403	1.222
		RMSE (ANN)								
		Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7	Hour 8	Average
Model 1	Day, Month, Hourly Volume Available Before the Missing Hours	0.392	0.647	0.922	1.004	0.761	0.401	0.333	0.291	0.594
Model 2	Day, Month, Hourly Volume Available Before the Missing Hours	0.474	0.651	0.907	1.012	0.768	0.496	0.355	0.365	0.629
Model 3(Monday)	Month, Hourly Volume Available Before the Missing Hours	0.240	0.358	0.455	0.385	0.341	0.435	0.457	0.409	0.385
Model 4 (January)	Day, Hourly Volume Available Before the Missing Hours	1.234	1.554	2.111	2.083	1.456	0.775	0.416	0.412	1.255

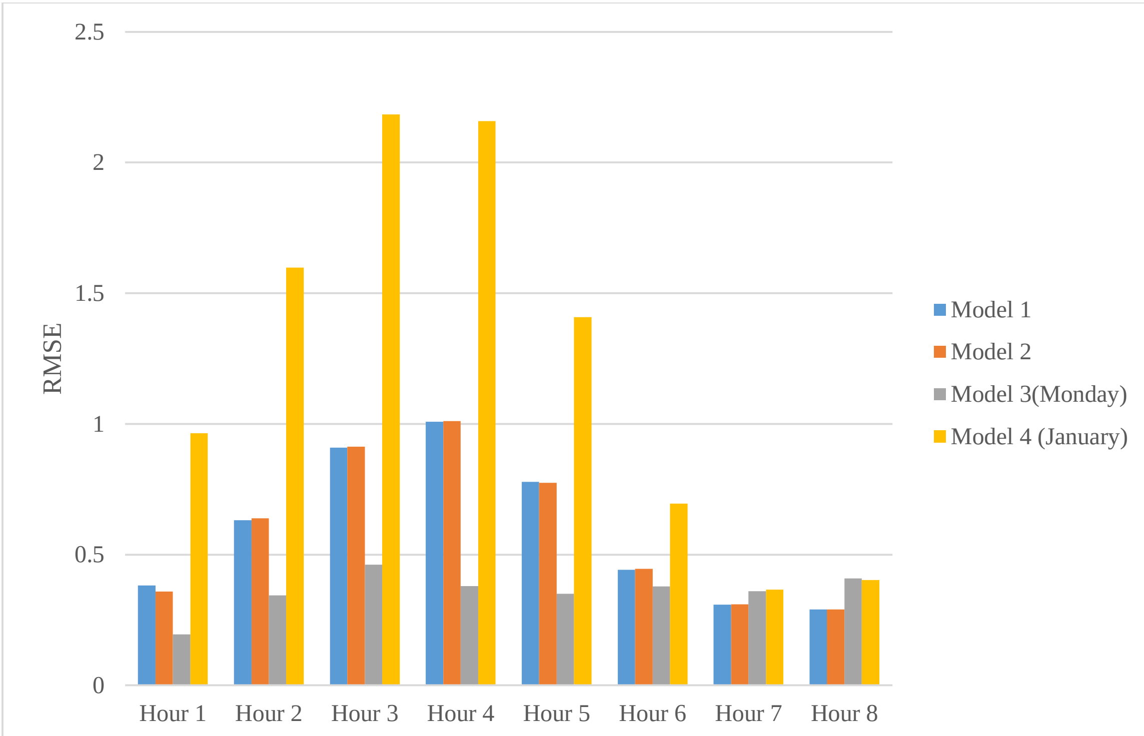

Figure 4-8: RMSE of Urban Principal Arterial – Interstate Model (SVR)

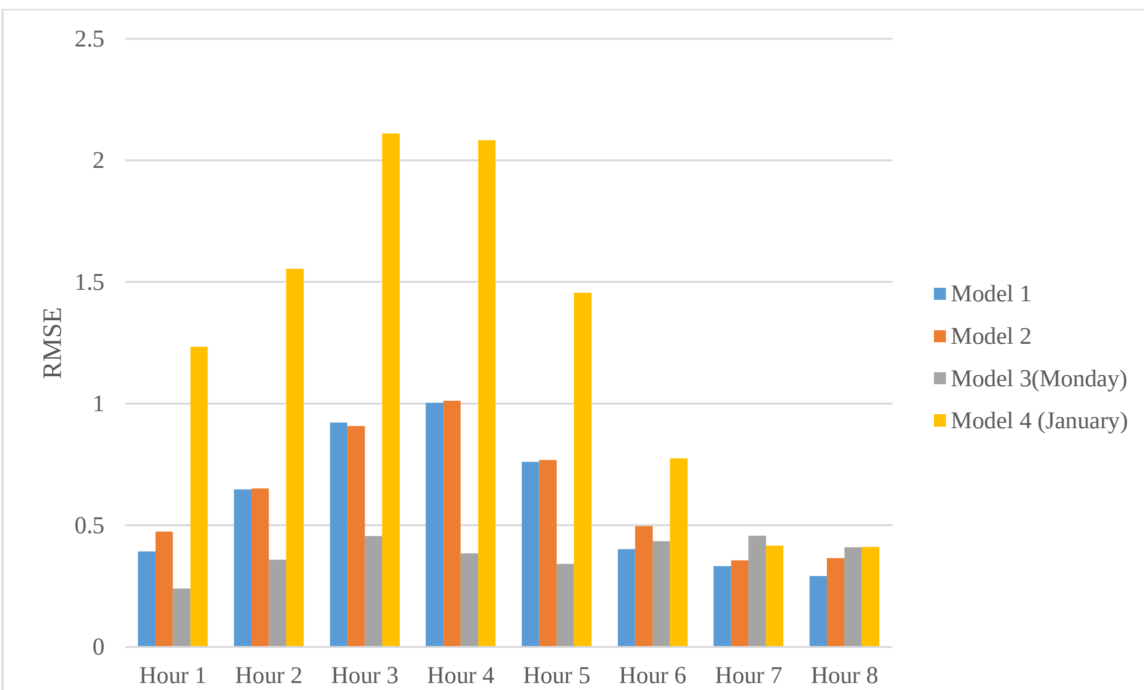

Figure 4-9: RMSE of Urban Principal Arterial – Interstate Model (ANN)

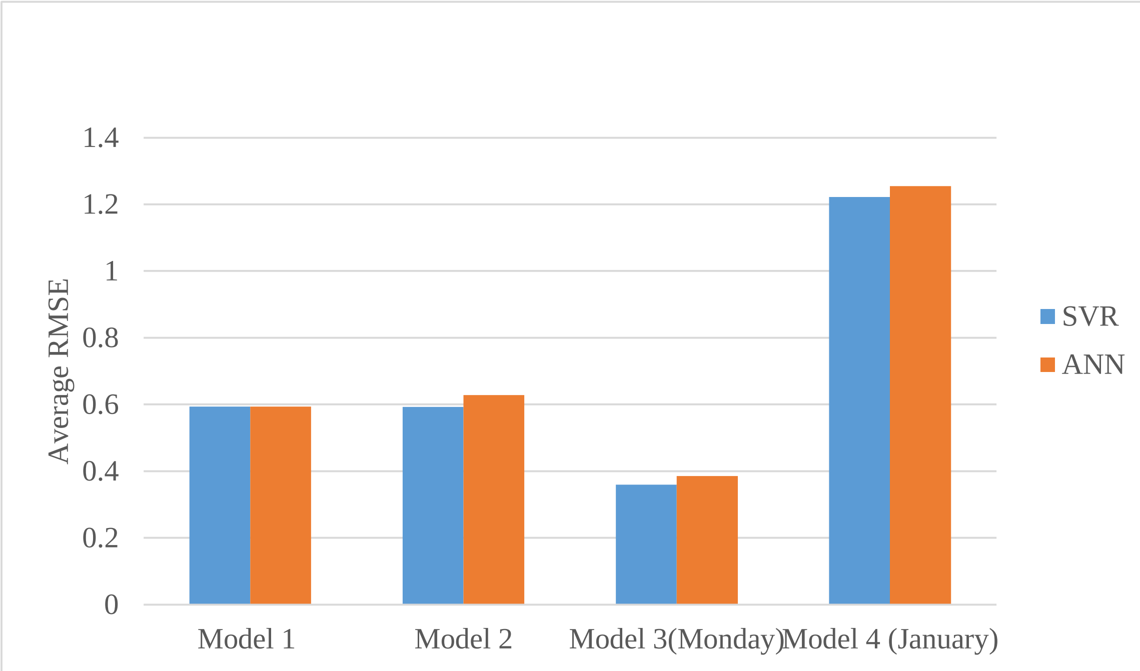

Figure 4-10: Average RMSE of Urban Principal Arterial – Interstate Model (SVR Vs ANN)

From the **Table 4-13** it can be concluded that AI Models developed using SVR for each of the hour performed better in predicting the hourly volume compared to the models developed using ANN for most of the hours. However, ANN predicted more accurately than SVR for some hours. In terms of the input features that developed least RMSE values are the month of the year categorical feature, direction of traffic, and the hourly volume. If the average RMSE values are compared for different models, it can be seen that the average RMSE values of SVR are less than the average RMSE values of the models developed using ANN. Please see appendices for the detailed RMSE calculation.

4.3.1.5 Model Evaluation: Rural Principal Arterial- Interstate

The rural principal interstate models were developed using 25 available permanent count stations. One of the characteristics of the ATRs used in this functional class having similar number of lanes. **Table 4-14** contains the RMSE values calculated for each of the models which are combination of different input features showed in **Table 4-7.** Figure 4-11 and 4-12 presents the graphical presentation of the errors for SVR and ANN and Figure 4-13 shows the graphical representation of the average RMSE of ANN and SVR.

Table 4-14: RMSE of Rural Principal Arterial – Interstate

	SVR									
	Input Features	Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7	Hour 8	Average
Model 1	Day, Month, Direction of Traffic, Hourly Volume Available Before the Missing Hours	0.536	0.573	0.605	0.629	0.654	0.547	0.576	0.622	0.593
Model 2	Day, Month, Hourly Volume Available Before the Missing Hours	0.532	0.570	0.602	0.629	0.654	0.548	0.456	0.483	0.559
Model 3	Month, Hourly Volume Available Before the Missing Hours	0.411	0.443	0.384	0.448	0.575	0.560	0.715	0.837	0.547
Model 4	Day, Hourly Volume Available Before the Missing Hours	0.436	0.470	0.511	0.509	0.561	0.512	0.561	0.631	0.524
	ANN									
	Input Features	Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7	Hour 8	Average
Model 1	Day, Month, Hourly Volume Available Before the Missing Hours	0.535	0.575	0.605	0.612	0.638	0.506	0.640	0.789	0.613
Model 2	Day, Month, Hourly Volume Available Before the Missing Hours	0.533	0.576	0.599	0.624	0.623	0.476	0.477	0.558	0.558
Model 3	Month, Hourly Volume Available Before the Missing Hours	0.412	0.563	0.494	0.542	0.646	0.563	0.761	0.881	0.608
Model 4	Day, Hourly Volume Available Before the Missing Hours	0.478	0.483	0.495	0.513	0.595	0.491	0.601	0.660	0.539

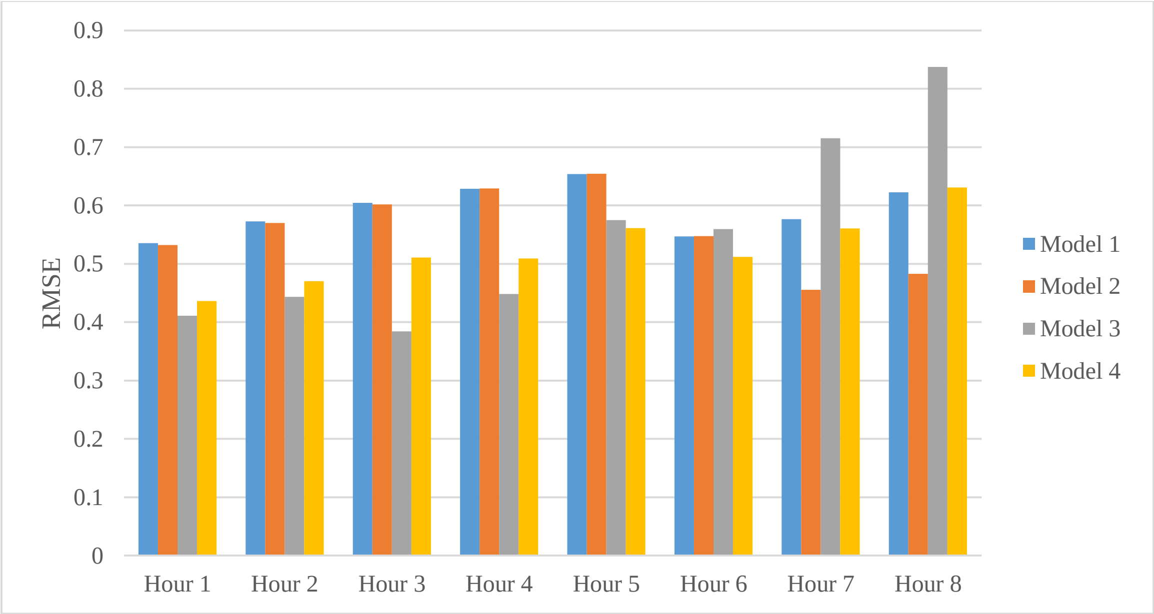

Figure 4-11: RMSE of Rural Principal Arterial – Interstate (SVR)

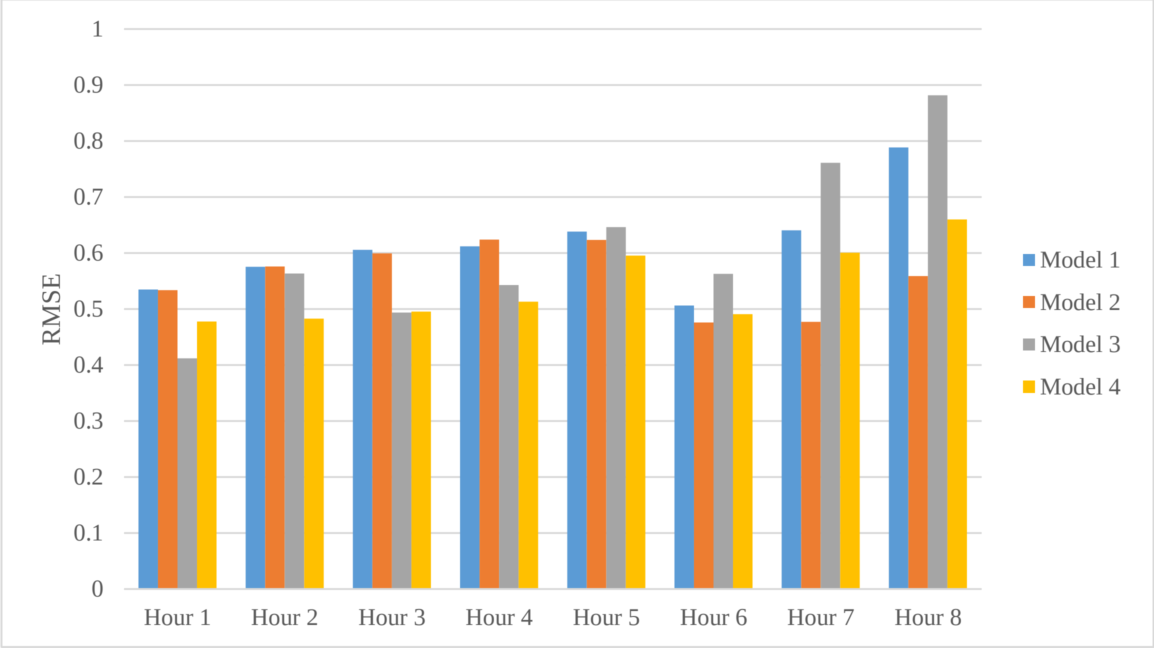

Figure 4-12: RMSE of Rural Principal Arterial – Interstate (ANN)

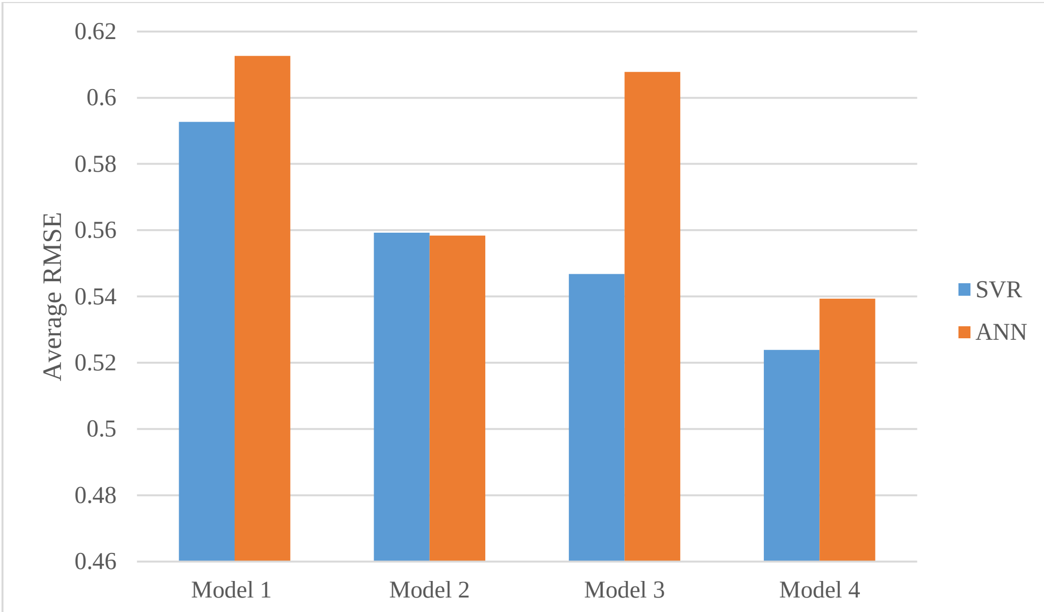

Figure 4-13: Average RMSE of Rural Principal Arterial – Interstate Model (SVR Vs ANN)

4.3.1.6 Urban Principal Arterial- Other

The urban principal arterial - other models were developed using 9 available permanent count stations. **Table 4-15** contains the RMSE values calculated for each of the models which are combination of different input features**.** The values of RMSE revealed the supremacy of SVR models over ANN models. Figures 4-14, 4-15 and 4-16 show the graphical representation of the errors.

Table 4-15 RMSE of Urban Principal Arterial – Other

		SVR								
		Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7	Hour 8	Average
Model 1	Day, Month, Hourly Volume Available Before the Missing Hours	0.325	0.424	0.500	0.535	0.460	0.445	0.455	0.479	0.453
Model 2	Day, Month, Hourly Volume Available Before the Missing Hours	0.311	0.383	0.456	0.531	0.393	0.267	0.229	0.218	0.349
Model 3	Month, Hourly Volume Available Before the Missing Hours	0.169	0.207	0.181	0.344	0.374	0.431	0.725	0.914	0.418
Model 4	Day, Hourly Volume Available Before the Missing Hours	0.307	0.543	0.546	0.431	0.310	0.444	0.375	0.491	0.431
		ANN								
		Hour1	**Hour2**	**Hour3**	**Hour4**	**Hour5**	**Hour6**	**Hour7**	**Hour8**	**Average**
Model 1	Day, Month, Hourly Volume Available Before the Missing Hours	0.414	0.499	0.997	0.952	0.490	0.555	0.600	0.634	0.643
Model 2	Day, Month, Hourly Volume Available Before the Missing Hours	0.669	0.449	0.433	0.543	0.461	0.341	0.265	0.250	0.426
Model 3	Month, Hourly Volume Available Before the Missing Hours	0.286	0.271	0.288	0.374	0.518	0.701	0.714	0.710	0.483
Model 4	Day, Hourly Volume Available Before the Missing Hours	0.891	0.643	0.594	0.460	0.346	0.362	0.421	0.778	0.562

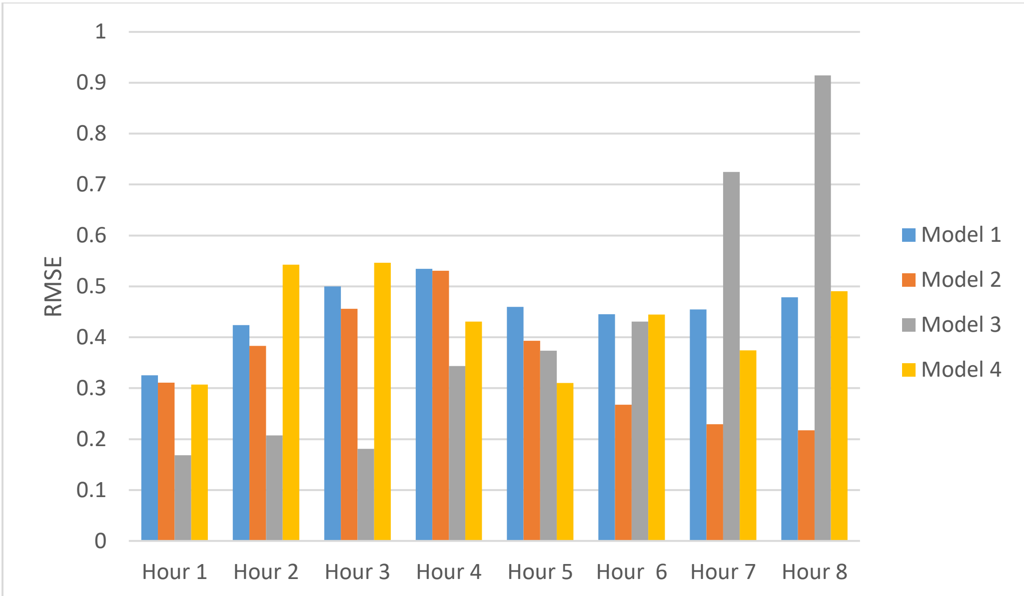

Figure 4-14: RMSE of urban Principal Arterial – other (SVR)

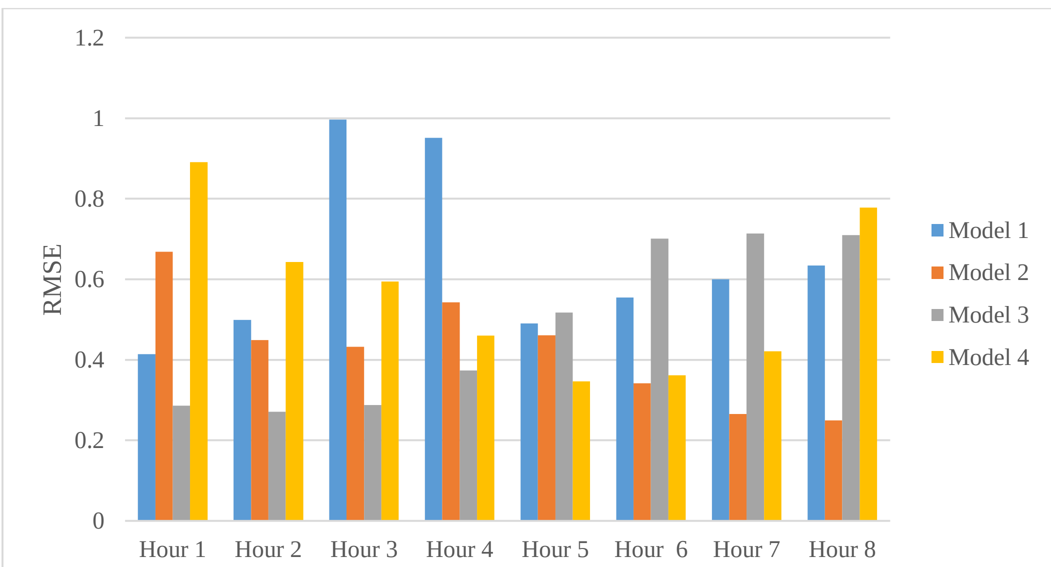

Figure 4-15: RMSE of urban Principal Arterial – other (ANN)

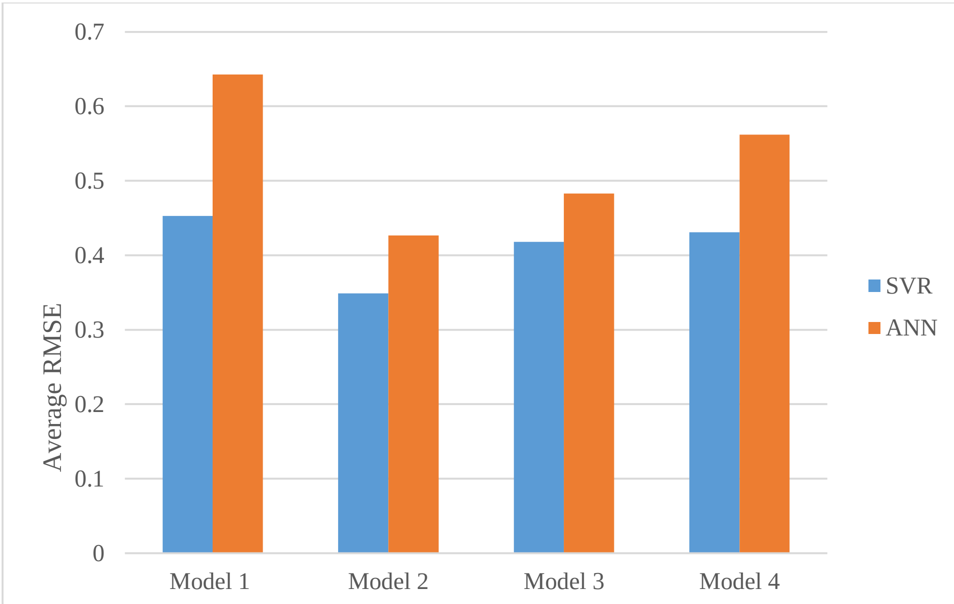

Figure 4-16: Average RMSE of Urban Principal Arterial – Other Models (SVR Vs ANN)

4.3.2 Comparison of Hourly Missing Volume Prediction using AI model and Historic Average Method

Currently, South Carolina DOT estimates missing hourly volume based on the historical average of the last three months of data for that particular hour and day. In this section of the study, a comparison was conducted between the accuracy of the prediction of missing hourly volume using SVR to the traditional method currently used by SCDOT for the Urban/Rural Principal Arterial functional class. In order to compare, 41 different days' data were randomly selected from different ATRs. The collected data were used for predicting the hourly volume using SVR. Once the hourly volume is predicted the values were compared with the current SCDOT method. **Figure 4-17** shows the Actual Vs Predicted Volume by SVR and historical average method by SCDOT

A paired t-test was conducted to determine if the differences between the actual hourly volume and the predicted volume for the hour 7AM-8AM with both of the methods is statistically significant. It was found that the difference between actual hourly volume and the predicted volume by SVR is not statistically significant at a 95% confidence level. However, there is a significant difference between the actual hourly volume and the predicted hourly volume using the historical average method practiced by SCDOT at a 95% confidence level. Thus, SCDOT could adopt the SVR model developed in this study to improve the missing value estimation accuracy.

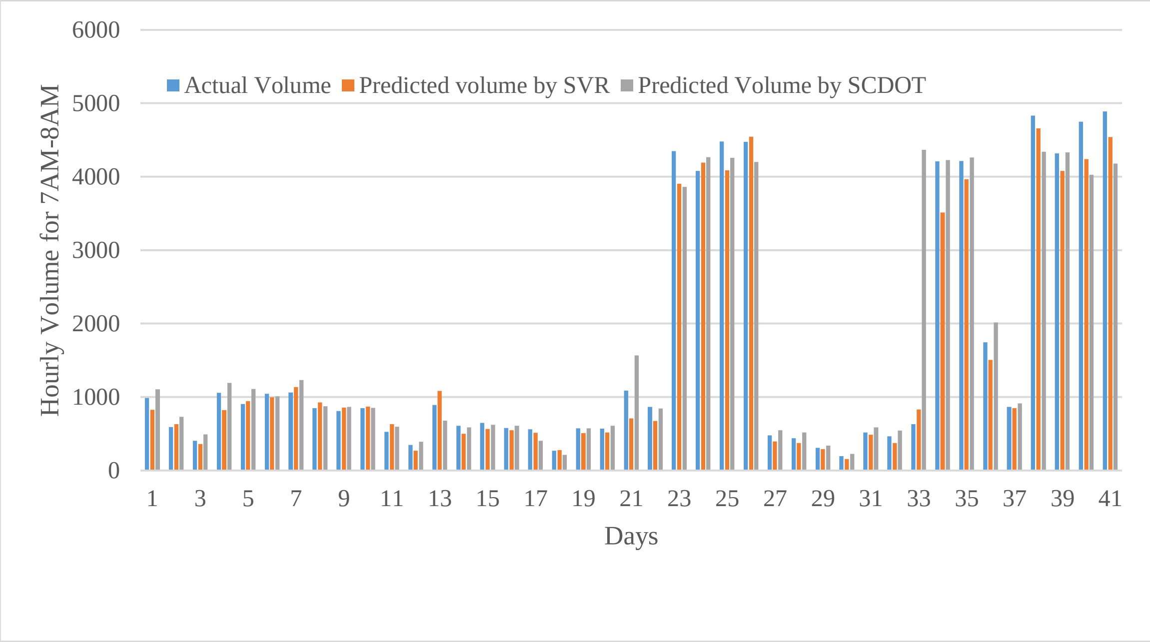

Figure 4-17: Actual verses Predicted Volume Estimated by SVR and historical average

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Overview

This chapter is divided in two sections. Section 5.2 presents conclusions based on the analysis conducted for this research. Following the conclusions, Section 5.3 presents recommendations of this research.

5.2 Conclusions

Average annual daily traffic (AADT) is one of the most important traffic information required for any traffic analysis. In this study, AADT estimation models for short-term count stations on different roadway functional classes in South Carolina were developed using Artificial Neural Network (ANN) and Support Vector Regression (SVR). This study revealed that AADT estimation models that use SVR outperformed the models that use ANN for Urban Principal Arterial-Interstate, Rural Principal Arterial-Interstate, Urban Principal Arterial-Other and Rural Principal Arterial-Other. The study revealed that the accuracy of estimation of AADT varies with different combinations of input features. In order to evaluate the AADT estimation models, the estimated AADTs for Urban Principal Arterial-Other and Rural Principal Arterial-Other functional classes were compared with the estimated AADT using factor method used by SCDOT. The results from the comparison showed that SVR produced lower MAPE and higher R^2 values than the traditional factor method. AADT estimation accuracy of the best performing SVR model was also compared with an OLS regression model for principal/minor arterial. This study revealed that the SVR model performed better than OLS regression model.

In addition to developing improved AADT estimation models, one other objective of this study was to solve the missing hourly volume problem at the permanent count stations operated and maintained by SCDOT. Transportation agencies often report that a significant portion of their hourly data collected from ATRs are missing or inaccurate. Although, currently SCDOT imputes the missing hourly volume using the average of the past three months' data for a particular hour, the method often produces unreliable estimations. In order to solve the aforementioned problem, this study developed models for imputing missing hourly volume using two Artificial Intelligence Paradigms (Artificial Neural Network, ANN and Support Vector Regression, SVR) that can be used for missing traffic data imputation for the roadways in South Carolina. The results from the analysis showed that the accuracy of the models varied based on the combination of the input features for different functional classes of roadways. However, this study revealed that missing hourly data estimation models using SVR performed better than the ANN models in terms of RMSE. Finally, it was found that AI based models outperform SCDOT's current historical average based missing value estimation method.

5.3 Recommendations

Based on the analysis conducted for this study, the following recommendations are made:

- This study revealed that SVR outperformed a regression-based model for estimating AADT. SVR should be further evaluated as a potential alternative to regression-based models for AADT estimation.

- In this study, SVR reliably imputed hourly volume that are missing at different permanent count stations. Therefore, SVR could potentially be applied for missing hourly volume imputation. However, follow-up studies are needed to establish the efficacies of SVR in missing volume imputation.
- There is a tradeoff between the AADT estimated methods currently used by state DOTs and SVR-based methods considered in this study. Therefore, it is recommended to estimate relative costs and benefits of these methods, which would aid in making an objective decision on suitable methods that can be adopted by state DOTs.

REFERENCES

Islam, S., Keehan, M., Chowdhury, M., Huynh, N., Torkjazi, M. and Shiri, S., "Study of Cost Effective Strategies for Estimating Statewide AADT", Project final report, 2017

Asif, M. T., N. Mitrovic, L. Garg, J. Dauwels, and P. Jaillet, Low-dimensional models for missing data imputation in road networks. In *2013 IEEE International Conference on Acoustics*, *Speech and Signal Processing*. IEEE, pp. 3527-3531, (2013).

Atluri, M., Chowdhury, M., Kanhere, N., Fries, R., Sarasua, W., and Ogle, J., "Development of a Sensor System for Traffic Data Collection," *Journal of Advanced Transportation*, Vol. 43, No. 1, pp 1-20, (2009).

Bhavsar, P., Chowdhury, M., Sadek, A., Sarasua, W., and Ogle, J., "Decision Support System for Predicting Traffic Diversion Impact across Transportation Networks using Support Vector Regression," *Transportation Research Record: Journal of the Transportation Research Board*, No. 2024, pp 100-106, (2007).

Castro, N., Manoel, Y. Jeong, M. K. Jeong, and L. D. Han. AADT prediction using support vector regression with data-dependent parameters. *Expert Systems with Applications* 36, pp. 2979-2986, (2009).

Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST) 2.3 (2011): 27.

Chowdhury, M., Sadek, A., Ma, Y., Kanhere, N., and Bhavsar, P., “Applications of Artificial Intelligence Paradigms to Decision Support in Real-time Traffic Management,” *Transportation Research Record: Journal of the Transportation Research Board*, No. 1968, pp 92-98, (2006).

Demuth, Howard, Mark Beale, and Martin Hagan. "Neural network toolbox™ 6." User Guide, COPYRIGHT 2008 (1992).

Drucker, H., D. Wu, and V.N., Vapnik, Support vector machines for spam categorization. IEEE Transactions on Neural networks, 10(5), pp.1048-1054. (1999).

Garber, Nicholas J., and Lester A. Hoel. *Traffic and highway engineering*. Cengage Learning, (2014).

Harwood, D.W., Bauer, K.M., Richard, K.R., G ilmore, D.K., Persaud, B., and Lyon, C. Development of SPFs for Safety Analyst Interim Tools: Technical Memorandum. 44 Prepared by Midwest Research Institute for the Federal Highway Administration, McLean, VA, (2004).

Himanen, V., P. Nijkamp, and A. Reggiani, Neural Networks in Transport Applications. Ashgate Publising, Farnham, Surrey, UK, (1998).

Jiang, Zhuojun, Mark R. McCord, and Prem K. Goel. "Improved AADT estimation by combining information in image-and ground-based traffic data."*Journal of transportation engineering* 132, no. 7: 523-530, (2006).

Keehan, M., Chowdhury, M., and Dey, K., “AADT Estimation with Regression Using Centrality and Roadway Characteristic Variables,” *Accepted for the 2017 Transportation Research Board Annual Meeting by the respective TRB committee*, Washington, D.C., (2017).

Kingan, Robert, and Todd Westhuis. "Robust regression methods for traffic growth forecasting." *Transportation Research Record: Journal of the Transportation Research Board* 1957: 51-55, (2006).

Langley, P. Selection of relevant features in machine learning. In *Proceedings of the AAAI Fall symposium on relevance, Vol. 184*, pp. 245-271, (1994).

Leverington, David, “A Basic Introduction to Feedforward Backpropagation Neural Networks”, (2009). Retrieved from http://www.webpages.ttu.edu/dleverin/neural_network/neural_networks.html (July 30, 2016).

Lin, W. A., Case Study on Support Vector Machine Versus Artificial Neural Networks. University of Pittsburgh, Pittsburgh, Pa., (2004).

Liu, Z., S. Sharma, and S. Datla. Imputation of missing traffic data during holiday periods. *Transportation Planning and Technology 31.5*, pp. 525-544 (2008).

Ma, Y., Chowdhury, M., Sadek, A., and Jeihani, M., "Integration of Vehicle Infrastructure Integration (VII) system and Artificial Intelligence (AI) for online travel time prediction," *IEEE Transactions on Intelligent Transportation Systems* Vol. 13, Issue: 3, pp 1369 – 1382, (2012).

Ma, Y., Chowdhury, M., Jeihani, M., and Fries, R., "Accelerated Incident Detection across Transportation Networks using Vehicle Kinetics and Support Vector Machine (SVM) in Cooperation with Infrastructure Agents," *IET ITS Journal*, Vol. 4, Issue 4, pp 328 -337, (2010).

Ma, Y., Chowdhury, M., Sadek, A., and Jeihani, M., "Real-Time Highway Traffic Condition Assessment Framework Using Vehicle-Infrastructure Integration (VII) with Artificial Intelligence (AI)," *IEEE Transactions on Intelligent Transportation Systems,* Vol. 10, No. 4, pp 615-627, (2009).

Mitchell, Tom M. "Machine learning." *Burr Ridge, IL: McGraw Hill* 45: 37, (1997).

Mohamad, D., Kumares S., Thomas K., and Charles S., "Annual average daily traffic prediction model for county roads."*Transportation Research Record: Journal of the Transportation Research Board* 1617: 69-77, (1998).

Muthukadan, B., Selenium with Python (2014). Retrieved from http://selenium-python.readthedocs.io/ (November 7, 2016)

Nelwamondo, F. V., S. Mohamed, and T. Marwala. Missing data: A comparison of neural network and expectation maximisation techniques. *arXiv preprint arXiv:0704.3474*, 2007.Chen, M., J. Xia, and R. Liu. Developing a strategy for imputing missing traffic volume data. In *Journal of the Transportation Research Forum*, *Vol. 45. No. 3*, (2010).

Performance Measurement System, PeMS, University of California, Berkeley. Available at http://pems.dot.ca.gov/, (July, 2016).

Qu, L., L. Li, Y. Zhang, and J. Hu. PPCA-based missing data imputation for traffic flow volume: a systematical approach. In *IEEE Transactions on intelligent transportation systems 10.3*, pp. 512-522 (2009).

Sharma, S., P. Lingras, F. Xu, G. Liu. Neural networks as alternative to traditional factor approach of annual average daily traffic estimation from traffic counts. Transportation Research Record: Journal of the Transportation Research Board 1660, pp. 24-31, (1999).

Sharma, S., P. Lingras, Fei Xu, and P. Kilburn. Application of neural networks to estimate AADT on low-volume roads. *Journal of Transportation Engineering* 127, no. 5, pp. 426-432, (2001).

Sharma, S., P. Lingras, and M. Zhong. Effect of missing value imputations on traffic parameters estimations from permanent traffic counts. In *Transportation Research Board*, (2003).

Traffic Monitoring Guide, FHWA -PL -95 -031, https://www.fhwa.dot.gov/ohim/tmgbook.pdf. (July 31, 2016).

Tu, Jack V. "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes." Journal of clinical epidemiology 49.11: 1225-1231.

Smith, J. ; U.S. Census Bureau; using American Fact Finder; http://factfinder2.census.gov. (January 24, 2016).

Vandervalk-Ostrander, A. *AASHTO Guidelines for Traffic Data Programs*. Aashto, (2009).

Van Lint, J. W. C., S. P. Hoogendoorn, and Henk J. van Zuylen. Accurate freeway travel time prediction with state-space neural networks under missing data. In *Transportation Research Part C: Emerging Technologies 13.5*, pp. 347-369, (2005).

Vanajakshi, Lelitha, and Laurence R. Rilett. "A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed." *Intelligent Vehicles Symposium, 2004 IEEE*. IEEE, (2004).

Web Scraping Retrieved from https://automatetheboringstuff.com/chapter11/ ((November 7, 2016)

Wu, C. H., J. M. Ho, and D. T. Lee. Travel-Time Prediction with Support Vector Regression. IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No. 4, pp. 276–281. 9, (2004).

Xia, Qing, Fang Zhao, Zhenmin Chen, L. Shen, and Diana Ospina. "Estimation of annual average daily traffic for nonstate roads in a Florida county." *Transportation Research Record: Journal of the Transportation Research Board* 1660: 32-40, (1999).

Yang, Bingduo, Sheng-Guo Wang, and Yuanlu Bao. "Efficient local AADT estimation via SCAD variable selection based on regression models." In *Chinese Control and Decision Conference (CCDC)*, pp. 1898-1902. IEEE, (2011).

Zhao, Fang, and Nokil Park. "Using geographically weighted regression models to estimate annual average daily traffic." *Transportation Research Record: Journal of the Transportation Research Board* 1879: 99-107, (2004).

Zhao, F. and Chung, S. "Estimation of annual average daily traffic in a Florida county using GIS and regression," presented at the Transportation Research Board Annual Meeting,Washington, DC, USA, Paper No. 01-3440, (2001).

Zhong, M., P. Lingras, and S. Sharma. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. In *Transportation Research Part C: Emerging Technologies 12.2*, pp. 139-166, (2004).

APPENDICES

APPENDIX A: MATLAB CODE FOR AADT ESTIMATION

Data Preparation Code

```
tic
ATR_ALL_FILE=zeros(0,0);
E_new=zeros(0,0);
for x=84:142%1:149%40:149%1:149%:33%:148%:100%:100%:0003
    Wednesday_AI=zeros(0,0);
    if exist (['ATR_' num2str(x) '_2011.txt'],'file')
           % if exist (['dta_i26_5% l' llos(lloss) '_' num2str(x)
'.str'],'file')
    %              delimiter = {',',';'};
    %              formatSpec = '%s%s%s%s%s%c%s%s%s%[^\n\r]';
    %              fileID = fopen(['dta_i26_5% l' llos(lloss) '_'
num2str(x)  '.fzp'],'r');
    %              dataArray = textscan(fileID, formatSpec, 'Delimiter',
delimiter,  'ReturnOnError', false);
    %              fclose(fileID);
    % toc
    % Import data from text file.
    % Initialize variables.
    % tic
    % filename =
'C:\Python34\matlab_AADT_Sababa\data_0002_jan_December.txt';
        delimiter = {',',' '};
        formatSpec = '%s%s%s%s%s%s%s%s%s%s%s%[^\n\r]';
        fileID = fopen(['ATR_' num2str(x) '_2011.txt'],'r');
        %fileID = fopen(filename,'r');
        dataArray = textscan(fileID, formatSpec, 'Delimiter',
delimiter,  'ReturnOnError', false);
        fclose(fileID);
        raw = repmat({''},length(dataArray{1}),length(dataArray)-1);
        for col=1:length(dataArray)-1
            raw(1:length(dataArray{col}),col) = dataArray{col};
        end
        numericData = NaN(size(dataArray{1},1),size(dataArray,2));

    for col=[1,2,3,4,5,6,7,8,9,10]
        % Converts strings in the input cell array to numbers. Replaced
non-numeric
        % strings with NaN.
        rawData = dataArray{col};
        for row=1:size(rawData, 1);
            % Create a regular expression to detect and remove non-
numeric prefixes and
            % suffixes.
            regexstr = '(?<prefix>.*?)(?<numbers>([-
]*(\d+[\,]*)+[\.]{0,1}\d*[eEdD]{0,1}[-+]*\d*[i]{0,1})|([-
]*(\d+[\,]*)*[\.]{1,1}\d+[eEdD]{0,1}[-+]*\d*[i]{0,1}))(?<suffix>.*)';
```

```
            try
                result = regexp(rawData{row}, regexstr, 'names');
                numbers = result.numbers;

                % Detected commas in non-thousand locations.
                invalidThousandsSeparator = false;
                if any(numbers==',');
                    thousandsRegExp = '^\d+?(\,\d{3})*\.{0,1}\d*$';
                    if isempty(regexp(thousandsRegExp, ',', 'once'));
                        numbers = NaN;
                        invalidThousandsSeparator = true;
                    end
                end
                % Convert numeric strings to numbers.
                if ~invalidThousandsSeparator;
                    numbers = textscan(strrep(numbers, ',', ''), '%f');
                    numericData(row, col) = numbers{1};
                    raw{row, col} = numbers{1};
                end
            catch me
            end
        end
    end

    % Replace non-numeric cells with NaN
    R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
    raw(R) = {NaN}; % Replace non-numeric cells
    % Create output variable
    five_d = cell2mat(raw);
    %Clear temporary variables
     clearvars filename delimiter formatSpec fileID dataArray ans raw
col numericData rawData row regexstr result numbers
invalidThousandsSeparator thousandsRegExp me R;
     toc
    % Import data from spreadsheet
    ATR_AADT =
xlsread('C:\Python34\matlab_AADT_Sababa\ATR_AADT_2011.xlsx','Sheet1');
    % Allocate imported array to column variable names
    VarName1 = ATR_AADT(:,1);
    VarName2 = ATR_AADT(:,2);
    % Clear temporary variables
     %%
    tic
    A=five_d;
    for ii=1:1:size(five_d,1)
           if five_d(ii,1)==999999;
            five_d(ii-5:ii,1)=999999;
           end
    end
    TF1 = (five_d(:,1)==999999);
    five_d(TF1,:) = [];
```

```
toc
%%
tic
fin=zeros(0,0);
for ii=1:29:(size(five_d,1)-29)
    ATR_num=repmat(five_d(ii,2),24,1);
    day=repmat(five_d(ii+1,2),24,1);
    month=repmat(five_d(ii+1,4),24,1);
    date=repmat(five_d(ii+1,5),24,1);
    year=repmat(five_d(ii+1,6),24,1);
    Fin=[ATR_num day month date year five_d((ii+5:ii+28),(1:8))];
    fin=[fin;Fin];
end
toc
%%
%AADT calculation from ATR using formula
Num_of_days = size(fin,1)/24;
new_fin= bsxfun(@plus,  fin(:,7), fin(:,10));
% new_fin = fin(:,7)+fin(:,10);
new_fin1 = [fin new_fin];
AADT_value=sum(new_fin1(:,14));
AADT=AADT_value/Num_of_days;
%%
%Insert AADT from SCDOT given value
Value = zeros(0,0);
for j=1:1:size(new_fin1,1)
    for k =1:1:size(ATR_AADT,1)
     if (new_fin1(j,1))==ATR_AADT(k,1)
         value=zeros(0,0);
         value=ATR_AADT(k,2);
         Value=[Value;value];
     else
         continue
     end
    end
end

%%
new_fin11=[new_fin1 Value];
%%
Wednesday=zeros(0,0);
for i=1:size(new_fin1,1)
    if new_fin11(i,2)==1
        wed1 = new_fin11(i,:);
        Wednesday=[Wednesday;wed1] ;
     else
        continue
    end
end
%%
tic
Wed_num=size(Wednesday,1)/24;
Wed_SADT=zeros(0,0);
```

```
    Wed_SADT3=zeros(0,0);
    for i=1:24:size(Wednesday,1)
        wed_SADT1=sum(Wednesday(i:i+23,14));
        wed_SADT2=repmat(wed_SADT1,24,1);
        Wed_SADT=[Wed_SADT;wed_SADT2] ;
    end

        Wed_SADT3 =[Wednesday Wed_SADT];
     toc

    %%
    tic
    %ADD additional parameters to the matrix that needs to be trained
    Add_parameters=zeros(0,0);
    for m=1:24:size(Wed_SADT3,1)
        Add_parameters1 = Wed_SADT3(m,1:4);
         Add_parameters=[Add_parameters;Add_parameters1] ;
      end
     toc
     %%
    tic
    Wed_final=zeros(0,0);
    % A = [1 2 10; 1 4 20;1 6 15] ;
    C = bsxfun(@rdivide,  Wed_SADT3(:,14), Wed_SADT3(:,16));
    D = bsxfun(@rdivide,  Wed_SADT3(:,15), Wed_SADT3(:,16));
    E=Wed_SADT3(:,1:16);
    toc
    %%
    Wed_trans=zeros(0,0);
    for i=1:24:size(C,1)
        Wed_trans2 = transpose (C(i:i+23));
        Wed_trans = [Wed_trans;Wed_trans2];
    end
    %%
    tic
    Actual_factor=zeros(0,0);
        for n=1:24:size(D,1)
        Actual_factor2 = D(n,1);
        Actual_factor=[Actual_factor;Actual_factor2] ;
        end
     toc
    %%
    %add 24 heading
    Wednesday_AI=[Add_parameters Wed_trans Actual_factor];
    end
    ATR_ALL_FILE=[ATR_ALL_FILE;Wednesday_AI];
    E_new=[E_new;E];

end
% Hour = [1:24];
% Wednesday_AI = vertcat(Hour,Wed_trans);
%%
```

```
% AAA=ATR_ALL_FILE;
ATR_ALL_Final=zeros(0,0);
tic
Jan=zeros(0,0);
ATR_ALL_FILE7=zeros(0,0);
    for pp=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(pp,3)==1;
            ATR_ALL_FILE7=ATR_ALL_FILE(pp,:);
            Jan=[Jan;ATR_ALL_FILE7];
           end
    end
 toc
tic
Feb=zeros(0,0);
ATR_ALL_FILE8=zeros(0,0);
    for pq=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(pq,3)==2;
            ATR_ALL_FILE8=ATR_ALL_FILE(pq,:);
            Feb=[Feb;ATR_ALL_FILE8];
           end
    end
 toc

 tic
Mar=zeros(0,0);
ATR_ALL_FILE9=zeros(0,0);
    for pr=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(pr,3)==3;
            ATR_ALL_FILE9=ATR_ALL_FILE(pr,:);
            Mar=[Mar;ATR_ALL_FILE9];
           end
    end
 toc

 tic
April=zeros(0,0);
ATR_ALL_FILE11=zeros(0,0);
    for zz=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(zz,3)==4;
            ATR_ALL_FILE11=ATR_ALL_FILE(zz,:);
            April=[April;ATR_ALL_FILE11];
           end
    end
 toc

 tic
May=zeros(0,0);
ATR_ALL_FILE12=zeros(0,0);
    for zk=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(zk,3)==5;
            ATR_ALL_FILE12=ATR_ALL_FILE(zk,:);
            May=[May;ATR_ALL_FILE12];
           end
```

```
    end
 toc
  tic
June=zeros(0,0);
ATR_ALL_FILE13=zeros(0,0);
    for zl=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(zl,3)==6;
            ATR_ALL_FILE13=ATR_ALL_FILE(zl,:);
            June=[June;ATR_ALL_FILE13];
           end
    end
 toc

  tic
July=zeros(0,0);
ATR_ALL_FILE14=zeros(0,0);
    for zm=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(zm,3)==7;
            ATR_ALL_FILE14=ATR_ALL_FILE(zm,:);
            July=[July;ATR_ALL_FILE14];
           end
    end
 toc
 tic
  August=zeros(0,0);
ATR_ALL_FILE15=zeros(0,0);
    for zn=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(zn,3)==8;
            ATR_ALL_FILE15=ATR_ALL_FILE(zn,:);
            August=[August;ATR_ALL_FILE15];
           end
    end
 toc
 tic
     September=zeros(0,0);
ATR_ALL_FILE16=zeros(0,0);
    for zo=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(zo,3)==9;
            ATR_ALL_FILE16=ATR_ALL_FILE(zo,:);
            September=[September;ATR_ALL_FILE16];
           end
    end
 toc
 tic
     October=zeros(0,0);
ATR_ALL_FILE17=zeros(0,0);
    for zp=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(zp,3)==10;
            ATR_ALL_FILE17=ATR_ALL_FILE(zp,:);
            October=[October;ATR_ALL_FILE17];
           end
    end
 toc
```

```
 tic
Nov=zeros(0,0);
ATR_ALL_FILE18=zeros(0,0);
    for ps=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(ps,3)==11;
            ATR_ALL_FILE18=ATR_ALL_FILE(ps,:);
            Nov=[Nov;ATR_ALL_FILE18];
           end
    end
 toc

  tic
Dec=zeros(0,0);
ATR_ALL_FILE19=zeros(0,0);
    for pt=1:1:size(ATR_ALL_FILE,1)
           if ATR_ALL_FILE(pt,3)==12;
            ATR_ALL_FILE19=ATR_ALL_FILE(pt,:);
            Dec=[Dec;ATR_ALL_FILE19];
           end
    end
 toc

  ATR_ALL_Final =
vertcat(Jan,Feb,Mar,April,May,June,July,August,September,October,Nov,De
c);
  %%
   tic
    %Test_Train=[other1 Train_final];
     fid4 = ['Thesis_monday_other_freeway_expressway_urban_AADT.xlsx'];
       xlswrite(fid4, ATR_ALL_Final);
       toc
     %%
     tic
     fid5= ['Thesis_monday_other_freeway_expressway_urban_AADT.xlsx'];

     xlswrite(fid5, E_new);
    toc

%%
```

Feature Selection Code:

```
%% Import data from spreadsheet
clear all; clc;
[~, ~, raw] = xlsread('Thesis_FC_13_data_AADT','Sheet1');
%[~, ~, raw] =
xlsread('data_imp_FC_1_11_24_12_normalize_data','24_hr_normalize_data
(3)');
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
```

```
raw(R) = {NaN}; % Replace non-numeric cells
SVMmonday = reshape([raw{:}],size(raw));
%%
X=SVMmonday(1:2596,34:57);
Y=SVMmonday(1:2596,58);

%%
% b = regress(Y,X);
% ds.Linear = b;
%%
opts = statset('display','iter');

fun = @(x0,y0,x1,y1) norm(y1-x1*(x0\y0))^2;  % residual sum of squares
[in,history] = sequentialfs(fun,X,Y,'cv',5, 'options',opts)
%%
```

Code for SVR

```
%% Import data from spreadsheet
[~, ~, raw] = xlsread('Thesis_FC_13_data_AADT','FC_1');
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};
% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
raw(R) = {NaN}; % Replace non-numeric cells
% Create output variable
SVMmonday = reshape([raw{:}],size(raw));
% Clear temporary variables
clearvars raw R;
%%
% test_f=SVMmonday(5248:7839,[9 13:50]);
% test_l=SVMmonday(5248:7839,51);
train_f=SVMmonday(2:276,[8:14 27:46]);
train_l=SVMmonday(2:276,52);
test_f=SVMmonday(277:414,[8:14 27:46]);
test_l=SVMmonday(277:414,52);

%% Run svr and get relative error
features_sparse = sparse(train_f); % features must be in a sparse
matrix
model=svmtrain(train_l,features_sparse,'-s 3 -t 2 -c 4096 -m 1  -g
0.01562 -d 1 -p .1 -e 0.00001');
model=svmtrain(train_l,features_sparse,'-s 3 -t 2 -c 2000 -g .5 -d 1 -p
.1 -e 0.00001');
features_sparse1 = sparse(test_f);
% model=svmtrain(train_l,features_sparse,'-s 3 -t 2 -c 20000 -g .000001
-d 3 -p .1 -e 0.00001');
% features_sparse1 = sparse(test_f);
[predict_label, accuracy, dec_values] =
svmpredict(test_l,features_sparse1,model);

Final=[test_l predict_label];
```

```
rel_err=bsxfun(@times, abs(bsxfun(@minus, Final(:,1), Final(:,2))),
100./(Final(:,1)));
avg=mean2(rel_err);
%%
% RMSE calculation
  actualandpredicted = bsxfun(@minus,  Final(:,1), Final(:,2));
   new2=bsxfun(@times,  actualandpredicted (:,1), actualandpredicted
(:,1));
   sum1=sum (new2(:,1));
   Y= size (new2,1) ;
  RMSE= sqrt(sum1/ Y);
  %%
```

APPENDIX B: MATLAB CODE FOR MISSING HOURLY VOLUME IMPUTATION

Data Preparation Code: 1

```
%% Import data from spreadsheet
[~, ~, raw] = xlsread('FC_6_12','Sheet2');
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};
% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
raw(R) = {NaN}; % Replace non-numeric cells
% Create output variable
SVMmonday = reshape([raw{:}],size(raw));
% Clear temporary variables
clearvars raw R;
%%
ATR_num=zeros(0,0);
ATR_num=unique(SVMmonday(:,1));
%%
tic
for i=1:1:8%(size(ATR_num,1))
    All=zeros(0,0);
    %fid4 = ['test' num2str(ATR_num(i,1)) '.xlsx'];
    fid4 = ['Other_freeway_espressway_82_142_' num2str(i) '.xlsx'];
    for j=1:1:(size(SVMmonday,1))
        if ATR_num(i,1)==SVMmonday(j,1)
            Single_ATR=SVMmonday(j,:);
            All=[All;Single_ATR];

        end
    end
    Alle=zeros(0,0);
    if sum(All(:,32))+ sum(All(:,33))==0
        [valuesN, orderN] = sort(All(:,30));
        North = All(orderN,:);
        Nrth_z=(North(:,30)==0);
        North(Nrth_z,:)=[];

        [valuesS, orderS] = sort(All(:,31));
        South = All(orderS,:);
        Soth_z=(South(:,31)==0);
        South(Soth_z,:)=[];

        Alle = [North; South];
    elseif sum(All(:,30))+sum(All(:,31))==0
        [valuesE, orderE] = sort(All(:,32));
        East = All(orderE,:);
        East_z=(East(:,32)==0);
        East(East_z,:)=[];

        [valuesW, orderW] = sort(All(:,33));
        West = All(orderW,:);
```

```
        West_z=(West(:,33)==0);
        West(West_z,:)=[];
        Alle = [East; West];
    end

    xlswrite(fid4,Alle);
end
toc
%%
```

Data Preparation Code: 2

```
tic
[~, ~, raw] =
xlsread('New_Urban_Rural_Principal_1_11_121_150_6','Sheet1');
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};
% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
raw(R) = {NaN}; % Replace non-numeric cells
% Create output variable
SVMmonday1 = reshape([raw{:}],size(raw));
% Clear temporary variables
clearvars raw R;toc
%%
tic
  All_hour=zeros(0,0);
  All_atr=zeros(0,0);
    for i=1:3:(size(SVMmonday1,1))
        Hour1= sum(SVMmonday1(i:i+2,34));
        Hour2= sum(SVMmonday1(i:i+2,35));
        Hour3= sum(SVMmonday1(i:i+2,36));
        Hour4= sum(SVMmonday1(i:i+2,37));
        Hour5= sum(SVMmonday1(i:i+2,38));
        Hour6= sum(SVMmonday1(i:i+2,39));
        Hour7= sum(SVMmonday1(i:i+2,40));
        Hour8= sum(SVMmonday1(i:i+2,41));
        Hour9= sum(SVMmonday1(i:i+2,42));
        Hour10= sum(SVMmonday1(i:i+2,43));
        Hour11= sum(SVMmonday1(i:i+2,44));
        Hour12= sum(SVMmonday1(i:i+2,45));
        Hour13= sum(SVMmonday1(i:i+2,46));
        Hour14= sum(SVMmonday1(i:i+2,47));
        Hour15= sum(SVMmonday1(i:i+2,48));
        Hour16= sum(SVMmonday1(i:i+2,49));
        Hour17= sum(SVMmonday1(i:i+2,50));
        Hour18= sum(SVMmonday1(i:i+2,51));
        Hour19= sum(SVMmonday1(i:i+2,52));
        Hour20= sum(SVMmonday1(i:i+2,53));
        Hour21= sum(SVMmonday1(i:i+2,54));
```

```
        Hour22= sum(SVMmonday1(i:i+2,55));
        Hour23= sum(SVMmonday1(i:i+2,56));
        Hour24= sum(SVMmonday1(i:i+2,57));
  all_hour=[Hour1 Hour2 Hour3 Hour4 Hour5 Hour6 Hour7 Hour8 Hour9
Hour10 Hour11 Hour12 Hour13 Hour14 Hour15 Hour16 Hour17 Hour18 Hour19
Hour20 Hour21 Hour22 Hour23 Hour24];
 All_atr=[All_atr; all_hour];
    end
      toc
     %%
     tic
 Date=zeros(0,0);
 Dir=zeros(0,0);
 All_date=zeros(0,0);
 final=zeros(0,0);
 % All_atr=zeros(0,0);
    for i=1:3:(size(SVMmonday1,1))
        ATR_num=SVMmonday1(i,1);
        Date=SVMmonday1(i,[2:5 7:8]);
        Day= SVMmonday1(i,9:15);
        Month=  SVMmonday1(i,16:27);
        Dir=  SVMmonday1(i,29:32);
%         Day= SVMmonday(i,2);
%         Month= SVMmonday(i,3);
%         Year=SVMmonday(i,4);
%         North=SVMmonday(i,5);
%         South=SVMmonday(i,6);
%         Dir=[North South ];
        Date_month_dir=[ATR_num Date Day Month Dir];
        final=[final;Date_month_dir];
           end
    toc
     %%
     tic
     final1=[final All_atr];
     toc
     %%
     %for 7-8AM only
  tic
  Train_final=zeros(0,0);
  Train_Lebel_Final=zeros(0,0);
  other1=zeros(0,0);
  ATR_num1=zeros(0,0);

%     A=five_d;
    for ii=2:1:size(final1,1)
           if (final1(ii,4)-final1(ii-1,4))==1
               train1= final1(ii-1,50:54);
               train2= final1(ii,31:38);
               train=[train1 train2];
               other=final1(ii,1:30);
               ATR_num=final1(ii,1);
               ATR_num1=[ATR_num1;ATR_num];
```

```
                Train_final=[Train_final;train];
                other1=[other1;other];

%                 train_lebel1=final((ii, 11);
%                 Train_Lebel_Final=[Train_Lebel_Final;train_lebel1];
            elseif (final1(ii,4)-final1(ii-1,4))==-30
                train1= final1(ii-1,50:54);
                train2= final1(ii,31:38);
                train=[train1 train2];
                other=final1(ii,1:30);
                ATR_num=final1(ii,1);
                ATR_num1=[ATR_num1;ATR_num];
                Train_final=[Train_final;train];
                other1=[other1;other];
            elseif (final1(ii,4)-final1(ii-1,4))==-29
                train1= final1(ii-1,50:54);
                train2= final1(ii,31:38);
                train=[train1 train2];
                other=final1(ii,1:30);
                ATR_num=final1(ii,1);
                ATR_num1=[ATR_num1;ATR_num];
                Train_final=[Train_final;train];
                other1=[other1;other];
               elseif (final1(ii,4)-final1(ii-1,4))==-27
                train1= final1(ii-1,50:54);
                train2= final1(ii,31:38);
                train=[train1 train2];
                other=final1(ii,1:30);
                ATR_num=final1(ii,1);
                ATR_num1=[ATR_num1;ATR_num];
                Train_final=[Train_final;train];
                other1=[other1;other];

            end
    end
    %%
%     Train_lebel = final1(2:361,33);
tic
    Test_Train=[other1 Train_final];
        fid4 = ['FC_1_11_ATR_138.xlsx'];
 xlswrite(fid4, Test_Train);
    toc
    %%
%  Date_Dir=final(2:364,1:5);
%  Test_Train1=[Date_Dir Test_Train];
%%
```

Feature Selection Code:

```
%% Import data from spreadsheet
clear all; clc;
```

```
[~, ~, raw] = xlsread('Thesis_FC_13_data_AADT','Sheet1');
%[~, ~, raw] =
xlsread('data_imp_FC_1_11_24_12_normalize_data','24_hr_normalize_data
(3)');
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
raw(R) = {NaN}; % Replace non-numeric cells
SVMmonday = reshape([raw{:}],size(raw));
%%
X=SVMmonday(1:2596,34:57);
Y=SVMmonday(1:2596,58);

%%
% b = regress(Y,X);
% ds.Linear = b;
%%
opts = statset('display','iter');

fun = @(x0,y0,x1,y1) norm(y1-x1*(x0\y0))^2;  % residual sum of squares
[in,history] = sequentialfs(fun,X,Y,'cv',5, 'options',opts)
%%
```

SVR Model Development Code:

```
%% Import data from spreadsheet
[~, ~, raw] = xlsread('thesis_FC_13_normalize_new_data','jan');
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x),raw)) = {''};
% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x),raw); % Find non-
numeric cells
raw(R) = {NaN}; % Replace non-numeric cells
% Create output variable
SVMmonday = reshape([raw{:}],size(raw));
% Clear temporary variables
clearvars raw R;
%%
%train_f=SVMmonday(2:1793,[5:11 34:44]);
train_f=SVMmonday(2:276,[8:14 27:46]);
train_l=SVMmonday(2:276,50);
test_f=SVMmonday(277:414,[8:14 27:46]);
test_l=SVMmonday(277:414,50);
%% Run svr and get relative error
tic
features_sparse = sparse(train_f); % features must be in a sparse
matrix
%model=svmtrain(train_l,features_sparse,'-s 3 -t 2 -c 32800 -m 1000  -g
.000075 -d 1 -p .1 -e 0.00001');
%model=svmtrain(train_l,features_sparse,'-s 3 -t 2 -c 20000 -m 1000  -g
.000005 -d 1 -p .1 -e 0.00001');
model=svmtrain(train_l,features_sparse,'-s 3 -t 2 -c 20000 -m 1000  -g
.0005 -d 1 -p .1 -e 0.00001');
```

```
features_sparse1 = sparse(test_f);

[predict_label, accuracy, dec_values] =
svmpredict(test_l,features_sparse1,model);
Final=[test_l predict_label];
rel_err=bsxfun(@times, abs(bsxfun(@minus, Final(:,1), Final(:,2))),
100./(Final(:,1)));
avg=mean2(rel_err);
toc
% RMSE calculation
  actualandpredicted = bsxfun(@minus,  Final(:,1), Final(:,2));
   new2=bsxfun(@times,  actualandpredicted (:,1), actualandpredicted
(:,1));
   sum1=sum (new2(:,1));
   Y= size (new2,1) ;
  RMSE= sqrt(sum1/ Y);
  %%
  result = [ RMSE avg];
  fid4 = ['updted_thesis_FC_13_dataimp_SVR_mode4_hr2.xlsx'];
    xlswrite(fid4, result);
   fid5 =
['updated_thesis_FC_13_dataimp_Actual_predicted_model4_hr2.xlsx'];
    xlswrite(fid5, Final);
```

APPENDIX C: RMSE CALCULATION FOR AADT ESTIMATION AND MISSING HOURLY VOLUME IMPUTATION

RMSE Calculation: AADT Estimation

RMSE calculated of Urban Principal Arterial- Interstate for model5 developed using ANN

No	Actual AADT Factor	Estimated AADT Factor	(Actual –Estimated)	(Actual-Estimated)^2
1	1.11282	1.084011	0.028809	0.00083
2	5.628326	7.754853	-2.12653	4.522116
3	1.097585	1.129151	-0.03157	0.000996
4	1.32908	1.139477	0.189603	0.035949
5	1.325301	1.082652	0.24265	0.058879
6	1.274269	1.017116	0.257153	0.066127
7	3.061338	2.21194	0.849398	0.721477
8	1.312837	1.066576	0.246261	0.060644
9	1.410835	1.075722	0.335113	0.112301
10	1.274682	1.086378	0.188304	0.035458
11	1.584923	1.194632	0.390291	0.152327
12	1.332816	1.101911	0.230905	0.053317
13	1.36551	1.102771	0.262739	0.069032
14	1.256913	1.107585	0.149328	0.022299
15	1.284567	1.229646	0.054921	0.003016
16	1.23004	1.128952	0.101088	0.010219
17	1.237011	1.196379	0.040633	0.001651
18	1.127558	0.946439	0.181118	0.032804
19	0.94975	0.928903	0.020847	0.000435
20	1.062473	0.926318	0.136156	0.018538
21	0.999201	0.956267	0.042934	0.001843
22	1.357019	1.441556	-0.08454	0.007146
23	1.421776	1.312396	0.10938	0.011964
24	1.400204	1.461855	-0.06165	0.003801
25	1.503883	1.368201	0.135682	0.01841
26	1.418476	1.418106	0.00037	1.37E-07
27	0.928301	1.183454	-0.25515	0.065103
28	1.335276	1.340253	-0.00498	2.48E-05

No	Actual AADT Factor	Estimated AADT Factor	(Actual –Estimated)	(Actual-Estimated)^2
29	1.505447	1.217365	0.288082	0.082991
30	1.436181	1.162678	0.273503	0.074804
31	1.241367	1.254649	-0.01328	0.000176
32	1.1557	1.161162	-0.00546	2.98E-05
33	6.911502	5.818268	1.093234	1.19516
34	1.166846	1.181334	-0.01449	0.00021
35	1.208796	1.138302	0.070494	0.004969
36	1.175093	1.095676	0.079417	0.006307
37	1.125369	1.199335	-0.07397	0.005471
38	4.108551	5.115213	-1.00666	1.01337
39	1.215519	1.225681	-0.01016	0.000103
40	1.245691	1.180911	0.06478	0.004196
41	1.135364	1.254381	-0.11902	0.014165
42	1.892326	1.511859	0.380467	0.144755
43	1.184027	1.129137	0.05489	0.003013
44	1.268608	1.112066	0.156542	0.024505
45	1.124295	1.120889	0.003406	1.16E-05
46	1.418551	1.275729	0.142822	0.020398
47	1.122339	1.104054	0.018285	0.000334
48	1.118402	1.099721	0.018681	0.000349
49	1.042147	0.914386	0.127762	0.016323
50	1.068612	0.912295	0.156317	0.024435
51	1.011383	0.887258	0.124125	0.015407
52	0.947652	0.888661	0.058991	0.00348
53	1.288084	1.996535	-0.70845	0.501903
54	1.438429	1.319184	0.119245	0.014219
55	1.408158	1.296608	0.11155	0.012443
56	1.43023	1.364201	0.066029	0.00436
57	1.312857	1.194148	0.118709	0.014092
58	1.050184	1.346335	-0.29615	0.087706
59	1.487377	1.193316	0.294061	0.086472
60	1.597595	1.270053	0.327542	0.107284
61	1.563492	1.205458	0.358034	0.128188
62	1.360032	1.204246	0.155786	0.024269
63	1.027727	1.440636	-0.41291	0.170493
64	5.796525	5.12157	0.674955	0.455564
65	1.028453	1.281431	-0.25298	0.063998
66	1.107984	1.3888	-0.28082	0.078857

No	Actual AADT Factor	Estimated AADT Factor	(Actual –Estimated)	(Actual-Estimated)^2
67	1.096915	1.350742	-0.25383	0.064428
68	1.074132	1.235409	-0.16128	0.02601
69	2.633482	1.580405	1.053077	1.108971
70	1.067624	1.341222	-0.2736	0.074856
71	1.131069	1.41638	-0.28531	0.081402
72	1.095073	1.354222	-0.25915	0.067158
73	1.291371	0.963623	0.327748	0.107419
74	1.072051	1.332343	-0.26029	0.067752
75	1.114767	1.321197	-0.20643	0.042613
76	1.058239	1.043635	0.014605	0.000213
77	1.092527	1.129161	-0.03663	0.001342
78	1.021052	1.117264	-0.09621	0.009257
79	1.04438	1.069558	-0.02518	0.000634
80	0.991087	1.208719	-0.21763	0.047364
81	0.923597	1.021776	-0.09818	0.009639
82	0.943111	1.244	-0.30089	0.090534
83	0.915107	1.066194	-0.15109	0.022827
84	1.368619	1.97205	-0.60343	0.36413
85	1.383262	1.680124	-0.29686	0.088127
86	1.312115	1.446029	-0.13391	0.017933
87	1.353276	1.394759	-0.04148	0.001721
88	1.289504	1.614004	-0.3245	0.1053
89	1.094388	1.343528	-0.24914	0.062071
90	1.446493	1.534849	-0.08836	0.007807
91	1.515257	1.572994	-0.05774	0.003334
92	1.498525	1.544262	-0.04574	0.002092
93	1.289923	1.575304	-0.28538	0.081443
94	1.028481	1.185154	-0.15667	0.024546
95	4.70627	5.545404	-0.83913	0.704146
96	1.125386	1.13682	-0.01143	0.000131
97	1.066031	1.212267	-0.14624	0.021385
98	0.983781	1.089469	-0.10569	0.01117
99	0.985867	1.258758	-0.27289	0.074469
100	2.27707	1.653348	0.623722	0.389029
101	1.038395	1.209233	-0.17084	0.029186
102	1.063494	1.310677	-0.24718	0.0611
103	1.007699	1.11513	-0.10743	0.011541
104	1.200647	1.067299	0.133348	0.017782

No	Actual AADT Factor	Estimated AADT Factor	(Actual –Estimated)	(Actual-Estimated)^2
105	1.018446	1.099034	-0.08059	0.006494
106	1.053407	1.131286	-0.07788	0.006065
107	0.981553	0.938232	0.043321	0.001877
108	1.013322	1.035143	-0.02182	0.000476
109	0.979167	0.975015	0.004152	1.72E-05
110	0.967376	0.979774	-0.0124	0.000154
111	0.898848	0.967511	-0.06866	0.004715
112	0.930581	0.993535	-0.06295	0.003963
113	0.866168	0.962404	-0.09624	0.009261
114	1.362132	1.609846	-0.24771	0.061362
115	1.354103	1.636035	-0.28193	0.079486
116	1.341306	1.686947	-0.34564	0.119468
117	1.42679	1.622295	-0.1955	0.038222
118	1.315547	1.499605	-0.18406	0.033877
119	1.17824	1.143601	0.034639	0.0012
120	1.435023	1.529819	-0.0948	0.008986
121	1.600179	1.437159	0.16302	0.026575
122	1.600985	1.350789	0.250196	0.062598
123	1.38529	1.294941	0.090349	0.008163
124	1.070019	1.123417	-0.0534	0.002851
125	6.48855	4.697864	1.790686	3.206555
126	1.080009	1.145828	-0.06582	0.004332
127	1.1047	1.112899	-0.0082	6.72E-05
128	1.084079	1.108123	-0.02404	0.000578
129	1.064961	1.04665	0.018311	0.000335
130	4.03278	5.020674	-0.98789	0.975934
131	1.091503	1.043545	0.047959	0.0023
132	1.130615	1.033917	0.096699	0.009351
133	1.076218	1.163499	-0.08728	0.007618
134	1.542522	1.527375	0.015147	0.000229
135	1.05859	1.116186	-0.0576	0.003317
136	1.121148	1.135	-0.01385	0.000192
137	1.049402	1.134621	-0.08522	0.007262
138	1.136801	1.269122	-0.13232	0.017509
139	1.029037	1.051712	-0.02267	0.000514
140	1.039755	1.110135	-0.07038	0.004953
141	0.959352	0.924316	0.035036	0.001227
142	0.930966	0.900145	0.030821	0.00095

No	Actual AADT Factor	Estimated AADT Factor	(Actual –Estimated)	(Actual-Estimated)^2
143	0.941496	0.922143	0.019353	0.000375
144	0.908899	0.928386	-0.01949	0.00038
145	1.365661	1.596083	-0.23042	0.053094
146	1.195984	1.236092	-0.04011	0.001609
147	1.140777	1.224316	-0.08354	0.006979
148	1.176224	1.141496	0.034728	0.001206
149	1.130593	1.143696	-0.0131	0.000172
150	1.21313	1.302404	-0.08927	0.00797
151	1.400574	1.329386	0.071188	0.005068
152	1.459683	1.3618	0.097883	0.009581
153	1.486171	1.31578	0.170391	0.029033
154	1.3554	1.275731	0.079668	0.006347
155	0.836193	1.240444	-0.40425	0.163419
156	2.92656	1.894318	1.032242	1.065524
157	1.032535	1.197292	-0.16476	0.027145
158	1.311475	1.126307	0.185168	0.034287
159	1.146411	1.179636	-0.03323	0.001104
160	1.050649	1.275138	-0.22449	0.050395
161	2.095172	2.451768	-0.3566	0.127161
162	1.282245	1.034061	0.248184	0.061595
163	1.377652	1.122312	0.25534	0.065199
164	1.105318	1.187696	-0.08238	0.006786
165	1.396069	1.545667	-0.1496	0.02238
166	1.309208	1.189862	0.119346	0.014243
167	1.364836	1.203041	0.161795	0.026178
168	1.073662	1.239279	-0.16562	0.027429
169	1.13795	1.354488	-0.21654	0.046889
170	1.180828	1.212838	-0.03201	0.001025
171	1.222395	1.427755	-0.20536	0.042173
172	0.980166	0.986507	-0.00634	4.02E-05
173	0.869565	0.920415	-0.05085	0.002586
174	1.049271	0.948074	0.101197	0.010241
175	0.9813	0.943586	0.037713	0.001422
176	0.795378	1.288148	-0.49277	0.242823
177	1.179088	1.493638	-0.31455	0.098942
178	1.200555	1.742942	-0.54239	0.294184
179	1.347144	1.378906	-0.03176	0.001009
180	1.23464	1.521506	-0.28687	0.082292

No	Actual AADT Factor	Estimated AADT Factor	(Actual –Estimated)	(Actual-Estimated)^2
181	0.594962	1.432647	-0.83768	0.701716
182	1.115525	1.456743	-0.34122	0.116429
183	1.286252	1.494848	-0.2086	0.043512
184	1.335097	1.326231	0.008865	7.86E-05
185	1.081246	1.3153	-0.23405	0.054781
186	1.066624	1.148365	-0.08174	0.006682
187	6.797874	5.861523	0.936351	0.876753
188	1.125974	1.216676	-0.0907	0.008227
189	1.044364	1.118207	-0.07384	0.005453
190	1.045872	1.129839	-0.08397	0.00705
191	1.015477	1.092955	-0.07748	0.006003
192	2.918467	3.011832	-0.09336	0.008717
193	1.017764	1.043804	-0.02604	0.000678
194	1.061106	1.125211	-0.0641	0.004109
195	1.014346	1.070076	-0.05573	0.003106
196	1.259852	1.03722	0.222633	0.049565
197	1.007308	1.041434	-0.03413	0.001165
198	1.041432	1.066006	-0.02457	0.000604
199	1.019802	0.953749	0.066053	0.004363
200	1.051343	0.982973	0.068369	0.004674
201	0.962925	0.947732	0.015193	0.000231
202	0.973342	0.99645	-0.02311	0.000534
203	0.904712	0.898441	0.006271	3.93E-05
204	0.915057	0.965445	-0.05039	0.002539
205	0.886601	0.955811	-0.06921	0.00479
206	1.577801	2.337353	-0.75955	0.576921
207	1.305585	1.482504	-0.17692	0.0313
208	1.277012	1.584786	-0.30777	0.094725
209	1.320879	1.521595	-0.20072	0.040287
210	1.266703	1.377768	-0.11107	0.012335
211	1.295761	1.296909	-0.00115	1.32E-06
212	1.487145	1.347274	0.139871	0.019564
213	1.581829	1.493225	0.088603	0.007851
214	1.546219	1.488572	0.057647	0.003323
215	1.388728	1.384842	0.003887	1.51E-05
				$\sum$= 24.44782

Total number of test cases = 215

RMSE (ANN) = √ (24.44782/215) = 0.33721

Missing Hourly Data Imputation

Accrual and predicted normalized hourly volume for the hour 12AM using model 3 using SVR.

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.4013697	0.50020137	-0.098831624	0.00976769
0.1147077	0.22012892	-0.105421197	0.011113629
0.0218454	0.21304347	-0.191198092	0.03655671
0.3488823	0.34251483	0.006367501	4.05451E-05
0.1510452	0.34798548	-0.196940323	0.038785491
0.1066327	0.0909153	0.015717434	0.000247038
0.0743328	0.17666294	-0.102330147	0.010471459
0.1800293	0.12911865	0.050910636	0.002591893
0.4377072	0.35028663	0.087420557	0.007642354
0.2850041	0.19597931	0.089024799	0.007925415
0.1759918	0.15460502	0.02138677	0.000457394
0.0750545	0.10389397	-0.028839504	0.000831717
0.0669795	0.1776488	-0.11066932	0.012247698
0.3617165	0.15060395	0.211112536	0.044568503
0.1954576	0.18222522	0.01323237	0.000175096
0.1840668	0.15192457	0.032142213	0.001033122
0.1679168	0.110892	0.057024805	0.003251828
0.1235044	0.25247807	-0.128973687	0.016634212
0.3132666	0.2275559	0.085710669	0.007346319
0.4505413	0.25777401	0.192767329	0.037159243
0.3173041	0.32048284	-0.00317878	1.01046E-05
0.0023796	0.1290239	-0.126644318	0.016038783

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.2648166	0.30827293	-0.043456287	0.001888449
0.2163667	0.2251267	-0.008759974	7.67371E-05
0.1025952	0.17837299	-0.075777743	0.005742266
0.0427545	0.1466181	-0.103863583	0.010787644
0.922928	0.97492787	-0.051999818	0.002703981
1.0036779	0.95861068	0.045067235	0.002031056
0.9350405	0.89814694	0.036893591	0.001361137
0.971378	0.95323478	0.018143187	0.000329175
0.7816158	0.78031433	0.001301453	1.69378E-06
0.9269655	0.93308899	-0.006123451	3.74967E-05
0.8381407	0.83513477	0.00300592	9.03555E-06
0.8179532	0.85287508	-0.034921856	0.001219536
0.7977658	0.80772981	-0.009964054	9.92824E-05
0.8098782	0.7469692	0.062909035	0.003957547
0.7654658	0.78375212	-0.018286305	0.000334389
0.7896908	0.81126145	-0.021570675	0.000465294
0.7049034	0.73288647	-0.027983055	0.000783051
0.8018033	0.7542729	0.04753035	0.002259134
0.7654658	0.68374769	0.081718121	0.006677851
0.660491	0.54000371	0.12048728	0.014517185
0.939078	0.75669394	0.182384087	0.033263955
0.644341	0.72430083	-0.079959815	0.006393572
0.7170159	0.59924081	0.117775086	0.013870971
0.7452783	0.74606424	-0.000785894	6.17629E-07
0.7291284	0.70221665	0.026911727	0.000724241
0.7614283	0.65052844	0.110899883	0.012298784
0.6160786	0.6034803	0.012598259	0.000158716
0.4909163	0.5152426	-0.024326328	0.00059177
0.660491	0.52425883	0.136232161	0.018559202
0.7049034	0.64889063	0.056012778	0.003137431
0.357679	0.50938819	-0.151709203	0.023015682
0.6806785	0.62085373	0.059824722	0.003578997
0.6039661	0.68698037	-0.083014291	0.006891372
0.676641	0.67463635	0.002004614	4.01848E-06
0.636266	0.55444241	0.081823613	0.006695104

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.7049034	0.69687307	0.00803034	6.44864E-05
0.7493158	0.74337544	0.0059404	3.52884E-05
0.8341032	0.73637443	0.097728771	0.009550913
0.7372034	0.63120496	0.105998403	0.011235661
0.947153	0.8098559	0.137297105	0.018850495
0.931003	0.79006929	0.140933748	0.019862321
0.9269655	0.7156762	0.211289344	0.044643187
0.8583282	0.80887967	0.049448484	0.002445153
0.7695033	0.53605867	0.233444635	0.054496398
0.7250909	0.72097019	0.004120692	1.69801E-05
0.8502532	0.67659954	0.173653633	0.030155584
0.7170159	0.53297583	0.184040064	0.033870745
0.8987031	0.87023249	0.028470596	0.000810575
0.8623657	0.89199153	-0.029625881	0.000877693
0.9269655	0.9030076	0.023957942	0.000573983
0.947153	0.87879599	0.068357024	0.004672683
0.6080036	0.53065418	0.077349399	0.005982929
0.8744781	0.84588408	0.028594046	0.000817619
0.9754155	0.86182459	0.113590874	0.012902887
0.8785156	0.79669257	0.081823053	0.006695012
0.660491	0.62158217	0.038908818	0.001513896
0.8785156	0.92208149	-0.043565862	0.001897984
0.971378	0.88099429	0.090383681	0.00816921
0.9915654	0.86804069	0.12352474	0.015258361
1.0359779	0.90013927	0.135838587	0.018452122
0.8704406	0.82938963	0.041051012	0.001685186
0.9592655	0.88479831	0.074467179	0.005545361
0.8987031	0.8210597	0.077643387	0.006028496
0.8462157	0.75430408	0.091911594	0.008447741
0.8825531	0.79864458	0.083908536	0.007040642
0.6564535	0.61987697	0.036576523	0.001337842
0.7977658	0.68419486	0.113570902	0.01289835
0.7331659	0.72012774	0.01303813	0.000169993
0.6726035	0.6006253	0.071978169	0.005180857
0.7695033	0.69062301	0.07888029	0.0062221

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.6564535	0.47881024	0.177643251	0.031557125
0.6241535	0.49827934	0.125874202	0.015844315
0.7291284	0.60230156	0.126826812	0.01608504
0.6887534	0.6804285	0.008324942	6.93047E-05
0.5635911	0.4715902	0.092000946	0.008464174
0.8219907	0.69296635	0.129024363	0.016647286
0.6483785	0.57132245	0.077056052	0.005937635
0.7250909	0.52564094	0.199449935	0.039780277
0.644341	0.59367421	0.0506668	0.002567125
0.5797411	0.53203513	0.047705996	0.002275862
0.5958911	0.45432394	0.141567156	0.02004126
0.5676286	0.57379966	-0.006171018	3.80815E-05
0.1759918	0.30928203	-0.133290241	0.017766288
0.5757036	0.58018888	-0.004485257	2.01175E-05
0.6120411	0.56605459	0.045986473	0.002114756
0.5474412	0.54264079	0.004800383	2.30437E-05
0.6080036	0.56911201	0.038891563	0.001512554
0.7049034	0.65263462	0.052268796	0.002732027
0.6322285	0.58999754	0.042230995	0.001783457
0.8300657	0.72647861	0.103587097	0.010730287
0.5312912	0.63569943	-0.104408232	0.010901079
0.8785156	0.78003276	0.098482865	0.009698875
0.7896908	0.77738001	0.012310764	0.000151555
0.7008659	0.61940397	0.081461953	0.00663605
0.7573908	0.76271884	-0.005328017	2.83878E-05
0.8219907	0.77161193	0.050378792	0.002538023
0.7129784	0.59405324	0.118925163	0.014143194
0.7210534	0.46375356	0.257299825	0.0662032
0.7937283	0.66061055	0.133117711	0.017720325
0.8421782	0.81242824	0.029749947	0.000885059
0.922928	0.83218818	0.090739865	0.008233723
0.971378	0.82972678	0.141651188	0.020065059
0.9996404	0.90655295	0.093087471	0.008665277
0.7533533	0.50937059	0.243982744	0.059527579
0.8260282	0.92624217	-0.100213955	0.010042837

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.9754155	0.87764343	0.097772028	0.00955937
0.7775783	0.78424634	-0.006668045	4.44628E-05
0.7291284	0.75808894	-0.028960571	0.000838715
0.8623657	0.6534455	0.208920152	0.04364763
0.7008659	0.59478497	0.106080945	0.011253167
0.8865906	0.61480599	0.271784616	0.073866878
0.6887534	0.45417177	0.23458167	0.05502856
0.7452783	0.54558804	0.199690305	0.039876218
0.7614283	0.60020725	0.161221065	0.025992232
0.6403035	0.43932663	0.200976891	0.040391711
0.8139157	0.54299413	0.2709216	0.073398513
0.7533533	0.55654633	0.196806997	0.038732994
0.7129784	0.46650603	0.246472369	0.060748629
0.4868788	0.30214704	0.184731734	0.034125814
0.6403035	0.40214517	0.238158353	0.056719401
0.7735408	0.44410948	0.32943132	0.108524994
0.7573908	0.48659238	0.270798449	0.0733318
0.4061289	0.20763559	0.198493321	0.039399598
0.8058407	0.54224139	0.263599354	0.06948462
0.5474412	0.36471156	0.182729618	0.033390113
0.6403035	0.3941558	0.24614772	0.0605887
0.5676286	0.42536881	0.142259828	0.020237859
0.5595537	0.36823549	0.191318169	0.036602642
0.5232162	0.32789389	0.195322321	0.038150809
0.5918536	0.41205108	0.179802525	0.032328948
0.5312912	0.3030618	0.228229406	0.052088662
0.652416	0.38686861	0.265547387	0.070515415
0.5716661	0.4663138	0.105352331	0.011099114
0.6201161	0.37874334	0.241372715	0.058260788
0.5393662	0.46951051	0.069855683	0.004879816
0.7452783	0.46382621	0.281452133	0.079215303
0.6322285	0.421048	0.211180529	0.044597216
0.7372034	0.53214256	0.205060801	0.042049932
0.4465038	0.34462266	0.101881187	0.010379776
0.7896908	0.67137077	0.118320003	0.013999623

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.8219907	0.62620245	0.195788269	0.038333046
0.6887534	0.46019005	0.228563388	0.052241222
0.8179532	0.52532341	0.292629809	0.085632205
0.8179532	0.53293474	0.285018481	0.081235534
0.7291284	0.61237496	0.116753413	0.013631359
0.6483785	0.48336093	0.165017578	0.027230801
0.7291284	0.59040954	0.13871883	0.019242914
0.8179532	0.67155107	0.146402153	0.02143359
0.7654658	0.60758558	0.157880231	0.024926167
0.7372034	0.68190464	0.055298717	0.003057948
0.7291284	0.71719784	0.011930534	0.000142338
0.7089409	0.44458642	0.264354487	0.069883295
0.7896908	0.64397972	0.145711053	0.021231711
0.8260282	0.51574172	0.310286493	0.096277708
0.636266	0.64484983	-0.008583805	7.36817E-05
0.660491	0.66858269	-0.008091708	6.54757E-05
0.7372034	0.74709114	-0.009887785	9.77683E-05
0.6726035	0.47729982	0.19530365	0.038143516
0.8139157	0.72774152	0.086174208	0.007425994
0.6120411	0.68017609	-0.068135027	0.004642382
0.8583282	0.64669767	0.211630483	0.044787461
0.7775783	0.60976557	0.167812721	0.028161109
0.6564535	0.59604751	0.060405981	0.003648883
0.7816158	0.67765722	0.103958564	0.010807383
0.7977658	0.63797754	0.159788213	0.025532273
0.7695033	0.62433167	0.145171635	0.021074803
0.5595537	0.51127444	0.048279211	0.002330882
0.7573908	0.59055776	0.166833066	0.027833272
0.8260282	0.67377609	0.152252122	0.023180709
0.4626538	0.50019554	-0.037541719	0.001409381
0.5676286	0.55137946	0.016249179	0.000264036
0.660491	0.60234345	0.058147537	0.003381136
0.4788038	0.46991573	0.008888059	7.89976E-05
0.6726035	0.5554799	0.117123562	0.013717929
0.7331659	0.50477653	0.228389335	0.052161689

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.5151412	0.48646946	0.028671767	0.00082207
0.652416	0.55775267	0.094663328	0.008961146
0.5797411	0.58600072	-0.006259599	3.91826E-05
0.5353287	0.57062549	-0.035296794	0.001245864
0.660491	0.62442746	0.036063529	0.001300578
0.5676286	0.5400517	0.027576943	0.000760488
0.6120411	0.54004914	0.071991927	0.005182838
0.5312912	0.55525857	-0.023967364	0.000574435
0.7816158	0.64835998	0.133255804	0.017757109
0.7008659	0.56599815	0.13486777	0.018189315
0.7452783	0.66818491	0.077093438	0.005943398
0.3455665	0.39311665	-0.047550139	0.002261016
0.8381407	0.72239664	0.115744048	0.013396685
0.8744781	0.74452643	0.129951704	0.016887445
0.7695033	0.68786706	0.081636248	0.006664477
0.7452783	0.6825488	0.062729544	0.003934996
0.8341032	0.64636744	0.187735753	0.035244713
0.7856533	0.6624655	0.12318778	0.015175229
0.7695033	0.73953894	0.029964367	0.000897863
0.6403035	0.58155816	0.058745361	0.003451017
0.8462157	0.72175672	0.124458959	0.015490033
0.8421782	0.76547841	0.076699777	0.005882856
0.8462157	0.72159936	0.124616315	0.015529226
0.7695033	0.66484032	0.104662981	0.01095434
0.7614283	0.61360781	0.147820505	0.021850902
0.7210534	0.63995977	0.081093617	0.006576175
0.7735408	0.58664783	0.186892972	0.034928983
0.6120411	0.47473089	0.13731018	0.018854085
0.5555162	0.65574453	-0.100228366	0.010045725
1.1894026	1.18462655	0.004776058	2.28107E-05
1.2055526	1.17353986	0.032012717	0.001024814
1.2378525	1.16738095	0.070471571	0.004966242
1.2297775	1.16501542	0.064762115	0.004194132
1.1934401	1.15104342	0.042396677	0.001797478
1.249965	1.17315792	0.076807082	0.005899328

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
1.1409527	1.14155987	-0.000607182	3.6867E-07
1.2136276	1.14531184	0.068315729	0.004667039
1.2095901	1.15185199	0.057738081	0.003333686
1.1530652	1.11960749	0.033457679	0.001119416
1.249965	1.15683768	0.093127321	0.008672698
1.2136276	1.14597481	0.067652751	0.004576895
1.1813276	1.12937066	0.051956957	0.002699525
1.2620775	1.16931596	0.092761522	0.0086047
1.1651776	1.09628828	0.068889364	0.004745744
1.22574	1.08620663	0.139533418	0.019469575
1.1611402	1.11721408	0.04392607	0.0019295
1.1732526	1.1250977	0.04815493	0.002318897
1.1894026	1.088727	0.100675606	0.010135578
1.1934401	1.13928986	0.054150236	0.002932248
1.1328777	1.10420269	0.028675012	0.000822256
1.0925028	1.13009164	-0.037588874	0.001412923
1.2136276	1.07728598	0.136341583	0.018589027
1.1692151	1.05815644	0.111058697	0.012334034
1.1894026	1.10401442	0.085388186	0.007291142
1.1409527	1.11294425	0.028008438	0.000784473
1.1772901	1.10969172	0.067598403	0.004569544
1.1853651	1.17049585	0.01486926	0.000221095
1.24189	1.13989299	0.101997023	0.010403393
1.1651776	1.1011749	0.064002748	0.004096352
1.1772901	1.13053635	0.046753776	0.002185916
1.2055526	1.11831933	0.087233247	0.007609639
1.1732526	1.14387889	0.029373745	0.000862817
1.249965	1.1287437	0.121221304	0.014694605
1.0763528	1.09227573	-0.015922938	0.00025354
1.249965	1.20262182	0.047343184	0.002241377
1.25804	1.1862288	0.071811194	0.005156848
1.2015151	1.17004353	0.031471553	0.000990459
1.1611402	1.14049644	0.020643715	0.000426163
1.2378525	1.135566	0.102286524	0.010462533
1.1894026	1.14095808	0.048444528	0.002346872

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
1.233815	1.13217426	0.101640775	0.010330847
1.1974776	1.144096	0.053381597	0.002849595
1.2136276	1.17098501	0.042642551	0.001818387
1.233815	1.16716359	0.06665144	0.004442414
1.266115	1.18030807	0.085806909	0.007362826
1.2540025	1.17715493	0.076847572	0.005905549
1.1571027	1.12531276	0.0317899	0.001010598
1.2176651	1.17865451	0.039010549	0.001521823
1.2701525	1.15354501	0.116607464	0.013597301
1.2217026	1.14664839	0.075054158	0.005633127
1.1894026	1.14673919	0.042663414	0.001820167
1.1934401	1.1415034	0.051936701	0.002697421
1.1409527	1.14555379	-0.004601107	2.11702E-05
1.1934401	1.14086524	0.052574856	0.002764115
1.2055526	1.10024791	0.105304668	0.011089073
1.1409527	1.10287047	0.038082213	0.001450255
1.233815	1.12163047	0.112184557	0.012585375
1.1853651	1.10974513	0.075619987	0.005718382
1.2217026	1.12047481	0.101227741	0.010247056
1.1611402	1.09343323	0.067706927	0.004584228
1.1126902	1.05708622	0.055604015	0.003091807
1.1853651	1.10407516	0.08128995	0.006608056
1.1651776	1.06338407	0.101793575	0.010361932
1.2095901	1.10393092	0.105659155	0.011163857
1.233815	1.0827931	0.151021935	0.022807625
1.1207652	1.06927148	0.051493738	0.002651605
1.1449902	1.04442508	0.100565099	0.010113339
1.1894026	1.06885566	0.120546943	0.014531565
1.0763528	0.73391549	0.342437303	0.117263307
1.1974776	1.043499	0.15397859	0.023709406
1.1611402	1.05807441	0.10306574	0.010622547
1.1046152	1.062157	0.042458243	0.001802702
1.2176651	1.06857907	0.149085991	0.022226633
1.0763528	1.05726322	0.019089573	0.000364412
1.1046152	0.99080769	0.113807553	0.012952159

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(Actual\text{-}Estimated)^2$
1.1248027	1.01695056	0.107852156	0.011632088
1.1248027	1.06285523	0.061947479	0.00383749
1.1207652	1.06807688	0.052688335	0.002776061
1.1571027	1.10224967	0.054852985	0.00300885
1.1853651	1.10145049	0.083914622	0.007041664
1.2015151	1.12645074	0.07506435	0.005634657
1.1934401	1.02093106	0.172509038	0.029759368
1.1853651	1.09356228	0.091802828	0.008427759
1.1772901	1.02676269	0.15052744	0.02265851
1.1853651	1.13143249	0.053932626	0.002908728
1.0359779	0.99013655	0.045841307	0.002101425
1.1611402	1.10317246	0.057967691	0.003360253
1.1813276	1.14541576	0.035911863	0.001289662
1.1692151	1.12754084	0.041674295	0.001736747
1.1732526	1.09374919	0.079503445	0.006320798
1.2055526	1.10577513	0.099777447	0.009955539
1.1490277	1.09943887	0.049588805	0.00245905
1.1853651	1.06265536	0.122709754	0.015057684
1.1369152	1.06876097	0.068154218	0.004644997
1.1611402	1.15433471	0.006805444	4.63141E-05
1.2136276	1.1058606	0.107766962	0.011613718
1.1772901	1.13922555	0.038064577	0.001448912
1.1369152	1.03192484	0.104990348	0.011022973
1.1530652	1.11101264	0.042052527	0.001768415
1.2015151	1.11067562	0.090839467	0.008251809
1.1692151	1.08121633	0.08799881	0.007743791
1.1611402	1.08338185	0.077758304	0.006046354
0.939078	0.99344423	-0.054366204	0.002955684
0.2519825	0.24078867	0.011193821	0.000125302
0.1308577	0.18579066	-0.05493297	0.003017631
0.3811823	0.48159457	-0.100412292	0.010082628
0.0056954	0.11631885	-0.110623442	0.012237546
0.6557318	0.73296682	-0.077235	0.005965245
0.1752701	0.36708674	-0.191816617	0.036793615
0.5144196	0.45548032	0.058939232	0.003473833

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(Actual-Estimated)^2$
0.4134822	0.38453661	0.028945614	0.000837849
0.4457822	0.38938013	0.056402038	0.00318119
0.6113194	0.69035916	-0.079039766	0.006247285
0.4054072	0.56781587	-0.162408634	0.026376564
0.4498197	0.53201453	-0.082194867	0.006755996
0.6476568	0.55958917	0.088067667	0.007755914
0.3892573	0.44864565	-0.059388382	0.00352698
0.6476568	0.68760636	-0.039949527	0.001595965
0.6516943	0.69498587	-0.04329154	0.001874157
0.5063446	0.56740467	-0.061060102	0.003728336
0.558832	0.72860552	-0.169773539	0.028823054
0.5063446	0.82717408	-0.320829512	0.102931576
0.5063446	0.4482909	0.058053673	0.003370229
0.6718818	0.62707645	0.044805345	0.002007519
0.4700071	0.68007956	-0.210072427	0.044130424
0.6153569	0.85383298	-0.238476097	0.056870849
0.8333815	0.84474016	-0.011358634	0.000129019
0.7162942	0.76078471	-0.044490488	0.001979404
1.0594811	0.71831539	0.341165759	0.116394075
1.0635186	0.83770004	0.225818598	0.050994039
0.7122567	0.76593066	-0.053673933	0.002880891
0.6516943	0.65155527	0.000139054	1.93361E-08
0.6839943	0.7754724	-0.091478129	0.008368248
1.0150687	0.88934912	0.125719599	0.015805418
0.6153569	0.64893409	-0.033577198	0.001127428
0.518457	0.6259741	-0.107517052	0.011559916
0.247945	0.42441837	-0.176473365	0.031142849
0.9343189	0.74956566	0.184753196	0.034133744
0.1954576	0.19311989	0.002337695	5.46482E-06
0.4094447	0.39289056	0.01655417	0.000274041
0.3488823	0.49946024	-0.150577911	0.022673707
0.5628695	0.38909378	0.173775697	0.030197993
0.4740446	0.45829678	0.015747839	0.000247994
0.4215572	0.5686779	-0.147120693	0.021644498
0.2116076	0.40694565	-0.195338093	0.038156971

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.4134822	0.3612638	0.052218426	0.002726764
0.2075701	0.25431553	-0.046745462	0.002185138
0.2762075	0.341307	-0.065099543	0.004237951
0.0581828	0.25627803	-0.198095215	0.039241714
0.2277575	0.19068391	0.037073626	0.001374454
0.6355444	0.69377295	-0.0582286	0.00339057
0.2519825	0.28002912	-0.028046623	0.000786613
0.3569573	0.37540483	-0.018447512	0.000340311
0.5467195	0.6341676	-0.087448093	0.007647169
0.566907	0.62553052	-0.058623554	0.003436721
1.0352562	0.64363363	0.391622557	0.153368227
0.8051191	0.64866216	0.156456915	0.024478766
0.9666188	0.74811274	0.218506058	0.047744897
0.7001442	0.44017159	0.259972662	0.067585785
1.6206927	1.37798233	0.242710379	0.058908328
0.9908438	0.76811737	0.222726397	0.049607048
1.0958186	0.85042531	0.245393273	0.060217858
1.0998561	0.6560947	0.443761381	0.196924164
0.9464313	0.82701003	0.119421307	0.014261448
1.467268	0.89131868	0.575949285	0.331717579
1.1200435	0.91456851	0.205475034	0.042219989
1.7862299	1.07569742	0.710532518	0.504856459
0.3932948	0.54178933	-0.148494576	0.022050639
0.22372	0.43554138	-0.211821342	0.044868281
0.4538572	0.62452807	-0.170670908	0.029128559
0.2963949	0.65503288	-0.358637958	0.128621185
0.2681325	0.45247241	-0.184339946	0.033981216
0.6557318	0.57293398	0.08279784	0.006855482
0.6113194	0.65200915	-0.040689753	0.001655656
0.27217	0.45438474	-0.182214779	0.033202226
1.1160061	0.82601716	0.28998889	0.084093556
0.6678443	0.48742859	0.180415715	0.03254983
0.7728191	0.69243757	0.080381551	0.006461194
0.9343189	1.03788628	-0.103567425	0.010726212
0.6759193	0.68616033	-0.010241043	0.000104879

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(\text{Actual-Estimated})^2$
1.0715936	0.6838318	0.387761824	0.150359232
1.0473687	1.03474092	0.012627746	0.00015946
0.829344	0.85418381	-0.02483978	0.000617015
0.6678443	0.79535553	-0.127511232	0.016259114
1.0514062	0.87681524	0.174590917	0.030481988
0.7728191	0.72911653	0.043702597	0.001909917
0.4821196	0.50716767	-0.025048058	0.000627405
0.4982696	0.59112693	-0.092857347	0.008622487
0.2681325	0.3565956	-0.088463132	0.007825726
0.7485942	0.78943928	-0.040845117	0.001668324
0.1793076	0.21778234	-0.038474729	0.001480305
0.22372	0.42927302	-0.205552982	0.042252028
0.4538572	0.47924793	-0.025390769	0.000644691
0.4175197	0.39207	0.025449721	0.000647688
0.0501078	0.36537865	-0.31527082	0.09939569
0.3690698	0.49897617	-0.129906373	0.016875666
0.0824078	0.40918287	-0.326775093	0.106781961
0.0662578	0.36265328	-0.296395475	0.087850278
0.0339579	0.26976573	-0.235807877	0.055605355
0.0218454	0.30482655	-0.28298117	0.080078343
0.0266045	0.25517494	-0.228570394	0.052244425
0.1631576	0.14095008	0.022207557	0.000493176
0.6355444	0.60369019	0.031854169	0.001014688
0.0501078	0.19798403	-0.147876204	0.021867372
0.3569573	0.32975411	0.027203209	0.000740015
0.4417447	0.59578009	-0.154035412	0.023726908
0.4700071	0.43728818	0.032718948	0.00107053
0.1638793	0.28765006	-0.123770751	0.015319199
0.5999286	0.62245633	-0.022527743	0.000507499
0.4020914	0.6243378	-0.222246385	0.049393456
0.628191	0.63978869	-0.011597651	0.000134505
0.5595537	0.65842672	-0.098873065	0.009775883
0.5595537	0.6348609	-0.075307242	0.005671181
0.4061289	0.4734507	-0.067321791	0.004532224
0.2204042	0.36803854	-0.14763432	0.021795892

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.3697915	0.48265787	-0.112866396	0.012738823
0.0185296	0.24931368	-0.230784122	0.053261311
0.1073544	0.21431751	-0.106963102	0.011441105
0.079092	0.19945661	-0.120364653	0.01448765
0.1396544	0.25678477	-0.117130415	0.013719534
0.2163667	0.35375611	-0.137389383	0.018875843
0.1881043	0.16861846	0.019485815	0.000379697
0.9100939	0.55704427	0.353049626	0.124644039
0.1961793	0.22821461	-0.032035348	0.001026264
0.3011541	0.30872607	-0.007571987	5.7335E-05
0.038717	0.26356506	-0.224848033	0.050556638
0.1558043	0.35768596	-0.201881633	0.040756194
0.0622203	0.21524401	-0.153023699	0.023416253
0.0992794	0.21668581	-0.117406393	0.013784261
0.0218454	0.20915218	-0.187306806	0.03508384
0.0056954	0.15569144	-0.149996035	0.02249881
0.0218454	0.1343177	-0.112472327	0.012650024
0.0379953	0.13843963	-0.100444276	0.010089053
1.2855808	0.71861164	0.566969136	0.321454001
0.0622203	0.13515574	-0.07293543	0.005319577
0.1227827	0.19502603	-0.072243324	0.005219098
0.1227827	0.20327119	-0.080488486	0.006478396
0.1833451	0.090431	0.092914111	0.008633032
0.0097329	0.10432504	-0.094592139	0.008947673
0.0864453	0.1082941	-0.02184883	0.000477371
0.325379	0.34207163	-0.016692589	0.000278643
0.1268202	0.2546903	-0.127870098	0.016350762
0.5393662	0.54831647	-0.008950285	8.01076E-05
0.4303539	0.49868751	-0.068333636	0.004669486
0.5111037	0.5734728	-0.062369064	0.0038899
0.4707288	0.56110676	-0.090377957	0.008168175
0.4586163	0.49753261	-0.038916291	0.001514478
0.1194669	0.34513515	-0.22566826	0.050926164
0.2728916	0.25909332	0.013798308	0.000190393
0.2163667	0.45364356	-0.237276829	0.056300293

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(Actual-Estimated)^2$
0.381904	0.49913149	-0.117227539	0.013742296
0.4020914	0.49925137	-0.097159958	0.009440057
0.5312912	0.59250877	-0.061217573	0.003747591
0.365754	0.47946074	-0.113706761	0.012929227
0.8818314	0.62813991	0.253691534	0.064359394
0.3092291	0.66626974	-0.357040665	0.127478036
0.4747663	0.55850204	-0.08373574	0.007011674
0.0871669	0.31002249	-0.222855545	0.049664594
0.5023071	0.19225745	0.310049626	0.096130771
0.4061289	0.45266692	-0.046538007	0.002165786
0.4465038	0.69643941	-0.249935563	0.062467786
0.6201161	0.68037578	-0.060259727	0.003631235
0.7089409	0.73582953	-0.026888621	0.000722998
0.7856533	0.70142693	0.084226348	0.007094078
0.7654658	0.76502338	0.000442433	1.95747E-07
0.3294165	0.46728857	-0.137872029	0.019008696
0.5555162	0.66402589	-0.108509725	0.011774361
0.357679	0.47216481	-0.114485822	0.013107003
0.2365542	0.4246737	-0.188119507	0.035388949
0.4061289	0.37981524	0.02631367	0.000692409
0.4949538	0.41233087	0.082622894	0.006826543
0.4545788	0.42141806	0.033160767	0.001099636
0.6160786	0.54663217	0.069446391	0.004822801
0.7445567	0.16585522	0.578701456	0.334895375
0.1961793	0.38466672	-0.188487459	0.035527522
0.325379	0.37970355	-0.054324502	0.002951152
0.4061289	0.4077552	-0.001626288	2.64481E-06
0.2890416	0.34878894	-0.059747331	0.003569744
0.4222789	0.45566265	-0.033383769	0.001114476
0.1477293	0.33158935	-0.183860012	0.033804504
0.4061289	0.40109724	0.00503167	2.53177E-05
0.2527042	0.26603394	-0.013329774	0.000177683
0.0912044	0.17744039	-0.086235951	0.007436639
0.046792	0.14856383	-0.10177182	0.010357503
0.0064171	0.06911736	-0.062700289	0.003931326

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(\text{Actual-Estimated})^2$
0.3044699	0.18998418	0.114485727	0.013106982
0.0137704	0.13539053	-0.121620139	0.014791458
0.0339579	0.22994934	-0.195991483	0.038412662
0.062942	0.07183598	-0.008893999	7.91032E-05
0.0622203	0.23099854	-0.168778229	0.028486091
0.0420328	0.20764624	-0.165613398	0.027427798
0.0097329	0.14458521	-0.134852315	0.018185147
0.5514787	0.4188253	0.132653364	0.017596915
0.0743328	0.32657829	-0.252245503	0.063627794
0.5757036	0.69323419	-0.117530565	0.013813434
0.5353287	0.63094673	-0.095618031	0.009142808
0.4545788	0.50326544	-0.048686613	0.002370386
0.4626538	0.55008311	-0.087429294	0.007643881
0.5878161	0.55423707	0.033579035	0.001127552
0.4989913	0.47654414	0.022447116	0.000503873
0.4626538	0.49337151	-0.030717698	0.000943577
0.4142039	0.46916931	-0.054965415	0.003021197
0.5070662	0.59609664	-0.089030398	0.007926412
0.6403035	0.60270614	0.037597383	0.001413563
0.6564535	0.71654513	-0.060091637	0.003611005
0.5353287	0.58382787	-0.04849917	0.00235217
0.1389327	0.07288321	0.066049471	0.004362533
0.5716661	0.62739014	-0.055724003	0.003105164
0.6080036	0.58192461	0.026078968	0.000680113
0.022567	0.19227502	-0.16970797	0.028800795
0.5547945	0.05718857	0.497605923	0.247611655
0.0864453	0.16875174	-0.082306474	0.006774356
0.0064171	0.08594302	-0.079525944	0.006324376
0.3044699	0.25540267	0.049067232	0.002407593
0.0945203	0.06230144	0.032218818	0.001038052
0.0501078	0.42410411	-0.373996276	0.139873215
0.1308577	0.19551246	-0.064654763	0.004180238
0.6032444	0.34444783	0.258796582	0.066975671
0.3488823	0.38357337	-0.034691041	0.001203468
0.1510452	0.3902116	-0.23916644	0.057200586

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.4942321	0.51580726	-0.021575169	0.000465488
0.3609948	0.32767554	0.033319269	0.001110174
0.5103821	0.40490488	0.105477179	0.011125435
0.6072819	0.4633993	0.143882606	0.020702204
0.3731073	0.29050675	0.082600543	0.00682285
0.7284067	0.43923634	0.289170355	0.083619494
0.6032444	0.73726438	-0.134019977	0.017961354
0.6920693	0.45719595	0.234873311	0.055165472
0.6839943	0.5797001	0.10429417	0.010877274
0.7445567	0.66580779	0.078748886	0.006201387
0.4861571	0.47078776	0.015369347	0.000236217
0.7970441	0.70729265	0.089751438	0.008055321
0.4861571	0.47011112	0.016045981	0.000257474
0.542682	0.62449518	-0.081813174	0.006693395
0.8333815	0.70899436	0.124387162	0.015472166
0.7324442	0.69835223	0.034091963	0.001162262
0.5305695	0.45901379	0.071555742	0.005120224
1.1886809	0.9216491	0.267031832	0.071305999
0.6274694	0.62142482	0.006044543	3.65365E-05
0.6234319	0.60457997	0.018851905	0.000355394
0.829344	0.69271721	0.136626825	0.018666889
0.6880318	0.72431055	-0.036278784	0.00131615
0.6920693	0.44261265	0.249456613	0.062228602
0.6557318	0.54417594	0.111555881	0.012444715
0.2802449	0.33934016	-0.059095214	0.003492244
0.7768566	0.72366357	0.053193045	0.0028295
0.3044699	0.18124196	0.123227948	0.015185127
0.3286949	0.29079885	0.037896019	0.001436108
0.4215572	0.35958846	0.061968748	0.003840126
0.3085074	0.3830804	-0.074573001	0.005561132
0.3246574	0.33080037	-0.006142998	3.77364E-05
0.4619321	0.51216735	-0.050235211	0.002523576
0.8091566	0.67285393	0.136302631	0.018578407
0.3488823	0.38868849	-0.039806156	0.00158453
0.2358325	0.27739678	-0.041564255	0.001727587

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
0.2802449	0.38174089	-0.101495947	0.010301427
0.1591201	0.12058117	0.038538981	0.001485253
0.2277575	0.20591767	0.021839861	0.00047698
0.6880318	0.75395983	-0.065928064	0.00434651
0.4094447	0.18347126	0.225973468	0.051064008
0.4134822	0.29989344	0.113588787	0.012902413
0.5790194	0.4332373	0.14578215	0.021252435
0.6880318	1.5032943	-0.815262537	0.664653004
0.534607	0.70604133	-0.171434306	0.029389721
0.3852198	0.41561562	-0.030395844	0.000923907
3.3325899	0.7158229	2.616766961	6.847469328
0.3811823	0.45816919	-0.076986913	0.005926985
0.5103821	0.59971391	-0.089331845	0.007980179
0.2681325	0.41547447	-0.147342001	0.021709665
2.7673408	1.77603395	0.991306849	0.982689269
0.9423938	1.87203061	-0.929636765	0.864224514
0.8858689	0.77343569	0.112433243	0.012641234
1.2613558	0.86638094	0.394974876	0.156005152
0.4740446	0.50506596	-0.031021334	0.000962323
0.5386445	0.68425404	-0.145609527	0.021202134
2.4403038	1.03515032	1.405153529	1.97445644
1.6651051	0.79999105	0.865114085	0.74842238
0.829344	0.6710345	0.158309528	0.025061907
0.9181689	0.90265551	0.01551337	0.000240665
0.2600575	0.57425593	-0.314198453	0.098720668
0.8656815	0.70893206	0.156749407	0.024570376
0.6718818	0.6786087	-0.006726901	4.52512E-05
1.1604185	0.89620739	0.264211091	0.0698075
0.9908438	0.71925556	0.271588205	0.073760153
0.6193944	0.55966703	0.059727356	0.003567357
0.9100939	0.75370847	0.156385431	0.024456403
1.124081	1.12017038	0.003910659	1.52933E-05
1.2613558	0.82648976	0.434866053	0.189108484
1.0796686	0.93869224	0.140976374	0.019874338
1.0958186	1.1387962	-0.042977611	0.001847075

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(Actual-Estimated)^2$
0.9141314	0.73658005	0.177551336	0.031524477
0.8172316	0.76734943	0.049882125	0.002488226
0.9868063	0.81170373	0.175102534	0.030660898
1.2775058	1.24063786	0.036867929	0.001359244
0.6678443	0.68289766	-0.015053361	0.000226604
1.0554437	0.73449452	0.320949133	0.103008346
0.8131941	0.52736882	0.285825235	0.081696065
0.7647441	0.74107681	0.023667324	0.000560142
0.2156451	0.28008441	-0.064439354	0.00415243
0.5992069	0.57796332	0.021243599	0.00045129
0.6032444	0.61868531	-0.015440902	0.000238421
0.4175197	0.54961843	-0.132098714	0.01745007
1.75393	1.38238599	0.371543998	0.138044943
0.7647441	0.59770497	0.167039172	0.027902085
1.3744056	0.73818775	0.636217869	0.404773176
0.6516943	0.58102903	0.070665299	0.004993585
0.4780821	0.40876903	0.069313083	0.004804304
0.7647441	0.71884422	0.045899921	0.002106803
0.3529198	0.75069391	-0.397774082	0.158224221
0.5992069	0.82888258	-0.22967566	0.052750909
2.9288405	1.58885766	1.339982876	1.795554108
0.3448448	0.47163379	-0.126788951	0.016075438
0.7808941	0.6927553	0.088138816	0.007768451
1.1523435	0.91875729	0.233586199	0.054562513
1.0473687	1.14926367	-0.101895003	0.010382592
1.2095901	1.15527559	0.054314483	0.002950063
1.1853651	1.17043079	0.014934318	0.000223034
1.1853651	1.14080833	0.04455678	0.001985307
1.2217026	1.16911021	0.052592342	0.002765954
1.1692151	1.11034197	0.058873165	0.00346605
1.2015151	1.1505175	0.05099759	0.002600754
1.1651776	1.13062814	0.034549509	0.001193669
1.1409527	1.11845495	0.022497738	0.000506148
1.1611402	1.11606808	0.045072072	0.002031492
1.1248027	1.09396573	0.030836978	0.000950919

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	(Actual-Estimated)^2
1.22574	1.13412518	0.091614866	0.008393284
1.1651776	1.12044222	0.044735429	0.002001259
1.1288402	1.10836064	0.020479564	0.000419413
1.1853651	1.12525045	0.060114658	0.003613772
1.1571027	1.0625381	0.094564561	0.008942456
1.1894026	1.01890397	0.170498633	0.029069784
1.1167277	1.06192521	0.054802514	0.003003316
1.1530652	1.0714658	0.081599367	0.006658457
1.0803903	1.02240374	0.057986542	0.003362439
1.1005778	1.10146181	-0.000884061	7.81563E-07
1.1288402	1.0485945	0.080245708	0.006439374
1.1005778	1.10595497	-0.005377221	2.89145E-05
1.1853651	1.04463393	0.140731182	0.019805266
1.1207652	1.04387815	0.07688707	0.005911621
1.1046152	1.06220303	0.042412216	0.001798796
1.0440528	1.05874976	-0.014696916	0.000215999
1.1530652	1.05536535	0.097699817	0.009545254
1.1288402	1.13924655	-0.010406343	0.000108292
1.1651776	1.11060558	0.054572069	0.002978111
1.1207652	1.02223204	0.098533176	0.009708787
1.1813276	1.06145935	0.11986827	0.014368402
1.1490277	1.09081217	0.058215507	0.003389045
1.1046152	1.10437819	0.000237057	5.61961E-08
1.1974776	1.08766744	0.109810149	0.012058269
1.0198279	0.99431102	0.025516868	0.000651111
1.2378525	1.14949388	0.088358642	0.00780725
1.2217026	1.11817919	0.103523362	0.010717086
1.1328777	1.12105114	0.011826556	0.000139867
1.1167277	1.11272884	0.003998883	1.59911E-05
1.1853651	1.09176248	0.093602635	0.008761453
1.1086527	1.08710327	0.021549474	0.00046438
1.1692151	1.08484454	0.084370598	0.007118398
1.1894026	1.08693455	0.102468058	0.010499703
1.1692151	1.13721578	0.031999355	0.001023959
1.1692151	1.11052277	0.058692369	0.003444794

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(\text{Actual-Estimated})^2$
1.2136276	1.14102031	0.072607258	0.005271814
1.1732526	1.14613279	0.027119841	0.000735486
1.0965403	1.04068666	0.055853597	0.003119624
1.1449902	1.14592427	-0.000934093	8.72529E-07
1.25804	1.10630846	0.151731531	0.023022457
1.1207652	1.08549197	0.035273249	0.001244202
1.1813276	1.06882548	0.112502135	0.01265673
1.1449902	1.13426374	0.010726435	0.000115056
1.1449902	1.08841912	0.056571059	0.003200285
1.1651776	1.08378972	0.081387921	0.006623994
1.1934401	1.07531495	0.118125146	0.01395355
1.1288402	1.06269944	0.066140764	0.004374601
1.1611402	1.07942488	0.081715274	0.006677386
1.1772901	1.07620085	0.101089278	0.010219042
1.1853651	1.05884333	0.126521783	0.016007762
1.1167277	1.02290123	0.093826501	0.008803412
1.0965403	1.03504235	0.061497913	0.003781993
1.1046152	1.09436785	0.010247397	0.000105009
1.1248027	1.01986399	0.104938727	0.011012136
1.1732526	1.07136868	0.101883955	0.01038034
1.1813276	1.04993903	0.131388586	0.01726296
1.1369152	1.02261344	0.114301748	0.01306489
1.1651776	1.03191649	0.133261154	0.017758535
1.1571027	1.04361799	0.113484674	0.012878771
1.0400154	0.79245574	0.247559617	0.061285764
1.1409527	0.97514585	0.165806831	0.027491905
1.1328777	1.00301642	0.129861284	0.016863953
1.0561653	0.99706565	0.059099677	0.003492772
1.1369152	1.01834853	0.118566658	0.014058052
1.0925028	1.0003614	0.09214137	0.008490032
1.0480903	0.9403973	0.107693043	0.011597792
1.0359779	0.97650585	0.059472007	0.00353692
1.0763528	0.99053495	0.085817839	0.007364701
1.0642403	0.99041421	0.073826107	0.005450294
1.1288402	1.0543492	0.074491003	0.00554891

Actual Normalized Volume	Estimated Normalized Volume	(Actual –Estimated)	$(\text{Actual-Estimated})^2$
1.1490277	1.06589242	0.083135248	0.006911469
1.1328777	1.0905438	0.042333899	0.001792159
1.1288402	0.96995148	0.158888721	0.025245626
1.1288402	1.04564375	0.083196458	0.006921651
1.1288402	1.01291014	0.115930065	0.01343978
1.2176651	1.0515955	0.166069555	0.027579097
1.0238654	0.94297956	0.080885817	0.006542515
1.1167277	1.05243957	0.064288159	0.004132967
1.1328777	1.10195244	0.030925259	0.000956372
1.1288402	1.07878592	0.050054285	0.002505431
1.1732526	1.0468988	0.12635383	0.01596529
1.1530652	1.05699344	0.096071731	0.009229777
1.1207652	1.05357623	0.067188994	0.004514361
1.1409527	1.05908008	0.081872609	0.006703124
1.1207652	1.03729278	0.083472443	0.006967649
1.1611402	1.10437111	0.056769042	0.003222724
1.1853651	1.06400823	0.121356886	0.014727494
1.1651776	1.09680432	0.068373322	0.004674911
1.1894026	1.01753618	0.171866429	0.029538069
1.0763528	1.0385489	0.037803889	0.001429134
1.1853651	1.07441347	0.110951646	0.012310268
1.1611402	1.05123909	0.109901059	0.012078243
1.1046152	1.01734662	0.087268623	0.007615812
0.931003	0.95684309	-0.025840058	0.000667709
			$\sum$=28.37154657

Total number of test cases = 747

RMSE (SVR) = $\sqrt{(28.37154657/747)}$ = 0.33721