

DEMAND FORECASTING USING CONVOLUTIONAL NEURAL NETWORKS
ON BIG REAL WORLD DATA

A Thesis
by
ISHAN KHATRI

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Aug 2017

Major Subject: Electrical Engineering

DEMAND FORECASTING USING CONVOLUTIONAL NEURAL NETWORKS
ON BIG REAL WORLD DATA

A Thesis

by

ISHAN KHATRI

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Lijun Qian
Chair of Committee

Lei Huang
Committee Member

Pamela Obiomon
Committee Member

Xiangfang Li
Committee Member

Pamela Obiomon
Head of Department

Kendall T. Harris
Dean, Roy G. Perry College of Engineering

Cajetan M. Akujuobi
Dean, Graduate Studies

Aug 2017

Major Subject: Electrical Engineering

ABSTRACT

Demand forecasting using convolutional neural networks on big real world data.

(Aug 2017)

Ishan Khatri, B.Eng., Lalbhai Dalpatbhai College of Engineering

Chair of Advisory Committee: Dr. Lijun Qian

Demand forecasting has become very significant in recent years because of its accurate and efficient resource allocation for companies. However, demand forecasting usually involves processing very large and sophisticated data sets and it is very challenging to achieve accurate prediction. Although there are many existing methods in the literature, they could not be directly applied to big data. In this thesis, a novel demand forecasting method using deep learning specifically, a Convolutional Neural Network (CNN) based model is proposed to process huge amounts of data. A unique feature of the proposed method is that it has adaptable computational intricacy even with the increasing dimensionality of the data. As a result, the proposed method has excellent scalability. Large data sets, namely the usage data of Bike Sharing Service in New York City are used to validate the proposed method. The goal is to predict hourly bike rental demand in every station of New York City to improve the service and environment, and eventually build a “Green” city. The results demonstrate the effectiveness and superior performance of the proposed scheme. This is also one of winning strategies of the student team from the CREDIT Center in the 2016 IEEE Big Data Analytics Competition organized by the IEEE Big Data Initiative.

ACKNOWLEDGEMENTS

First I would like to thank my Advisor Dr. Lijun Qian, of the Electrical and Computer Engineering department at Prairie View A&M University. The endless support and opportunity he provided me is unexplainable. The door to Dr.Lijun Qian's office was always open for me and whenever I faced problems and needed any advice he was always there to steer me to the right direction .

I would also like to thank Dr. Xishuang Dong, Post doc of Electrical engineering department, without his passionate participation and input, this thesis could not have been successfully conducted.

This research work is supported in part by the U.S. Office of the Assistant Secretary of Defense for Research and Engineering (OASD(R&E)) under agreement number FA8750-15-2-0119, and by the U.S. Army Research Office under agreement number W911NF-16-1-0496. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Office of the Assistant Secretary of Defense for Research and Engineering (OASD(R&E)), the Army Research Office, or the U.S. Government.

I would also like to thank all the members of the CREDIT (Center of excellence in Research and Education for big military Data InTelligence) center for their support, they have not only helped me with their technical knowledge they are also the most friendly and helpful people I met.

Finally I would like to thank my parents, my brothers, my sister-in-law, my uncle

and aunt who are the only family I have in Texas for their unbelievable contribution and providing unfailing support and encouragement throughout my years of study and the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	x
1. INTRODUCTION AND BACKGROUND	1
1.1 Overview	1
1.1.1 Demand Forecasting	2
1.2 Methodologies	3
1.2.1 Survey Methods for Demand Forecasting	4
1.2.2 Statistical Methods for Demand Forecasting	4
1.2.3 Big Data & its Challenges	9
1.3 Thesis Outline	11
2. LITERATURE REVIEW	13
2.1 Theory	13
2.1.1 The Basics	13
2.2 “Potential” and “Problem”	14
2.3 Challenges	15
2.3.1 Skills	15
2.3.2 Unwanted Data / Noise	16
2.3.3 Hardware and Software	16
2.3.4 Algorithm Architecture	17
2.3.5 Statistical Importance	17
2.4 Applications of Demand Forecasting	17
2.4.1 Energy Demand Forecasting	18
2.4.2 Health Care Demand Forecasting	18
2.4.3 Transportation Demand Forecasting	19
2.5 Transportation Demand Forecasting : Bike Share Service Approach	19
2.5.1 Related Works	21

3. METHODOLOGY	24
3.1 The Problem	24
3.2 Datasets	25
3.2.1 New York City Bike Dataset	26
3.2.2 New York City Weather Dataset	27
3.2.3 New York City Holiday Dataset	27
3.3 Statistical Analysis	28
3.3.1 Weekly Distribution	29
3.3.2 Trip Duration Frequency	32
3.3.3 Hourly Distribution	32
3.4 Methodology	35
3.4.1 Convolutional Neural Network	36
3.4.2 Problem Formulation	38
3.4.3 Model Structure	40
4. EXPERIMENT & RESULTS	43
4.1 Dataset Combination & Feature Selection	43
4.2 Baseline Algorithms	43
4.2.1 Linear Regression	44
4.2.2 Neural Network	46
4.2.3 Support Vector Regression	46
4.3 Evaluation Metrics	48
4.4 Results	49
4.5 Discussion	50
5. CONCLUSION & FUTURE WORK	52
5.1 Conclusion	52
5.2 Future Work	53
REFERENCES	55
VITA	62

LIST OF FIGURES

FIGURE	Page
1.1 Architecture of the traditional Artificial Neural Network (ANN) (above) and the Convolutional Neural Network (CNN) (below). . . .	8
1.2 A sample Convolutional neural network architecture [1]	9
1.3 5 V's of Big Data	10
3.1 Dataset Overview.	25
3.2 Holiday list of NYC in 2015 (Jan-July) [2]	29
3.3 Customer vs Subscribers rental trend	30
3.4 Male vs Female average rental demand per week	31
3.5 Gender distribution in rental demand of bikes	31
3.6 Trip duration distribution between subscribers and customers	33
3.7 Hourly average rental demand in the City Bike trip data (Monday-Friday of Jan-June'2015)	34
3.8 Hourly average rental demand in the City Bike trip data (Saturday-Sunday of Jan-June'2015)	35
3.9 Comparing Image pixel with Bike rental demand sample	39
3.10 Structure of a CNN based regression model	41
3.11 5 Layer CNN	41
3.12 6 Layer CNN	41
3.13 7 Layer CNN	42
4.1 Simple Linear Regression [3]	45
4.2 Residual example [3]	45

4.3	Abstract neuron [4]	46
4.4	Functional model of a Neural Network [4]	47
4.5	Non-linear SVR [5]	48

LIST OF TABLES

TABLE	Page
3.1 Details of the Citi Bike Trip data	26
3.2 Target dataset of the Citi Bike trip data	27
3.3 Target weather dataset	28
3.4 Target Holiday dataset	28
4.1 Feature selection	44
4.2 Dataset counts	50
4.3 Comparison of results between proposed and baseline models (Female dataset)	51
4.4 Comparison of results between proposed and baseline models (Male dataset)	51

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 Overview

Predicting or forecasting demand for a product or service is now a common exercise for companies before or after launching a product in order to improve it for the customers. This type of forecasting is possible because of the availability of the “Big Data”. Big Data delivers many potentially breathtaking opportunities for researchers for discovering knowledge which is proving to be very valuable for any companies growth in the market share. However, forecasting using Big Data is not something which is free from challenges. It raises few questions which is also needed to be answered theoretically and methodically. Suppose, the significance of the data that can be found today, might show representation of whats happening currently (Now casting) or a very good tool to study for predicting what might happen in future (Forecasting). Methodically, there is also a need to prevent picturing unreliable conclusions also most importantly finding ways to reduce the dimensionality of the current Big Data and use it to find important knowledge.

Accurately demand forecasting paves new ways for improvement in many real world applications [6]. It helps organizations to reduce costs, take smart decisions and increase efficiency. From health care to energy to transportation, demand forecasting is helping researchers or decision makers to make better selections and crack problems which has a long lasting effect. The foundation of this forecasting technique came from the century old disciplines of statistics and mathematics. Continuous im-

This thesis follows the style of the IEEE transaction on Neural Networks and Learning Systems.

provement of analytics and mathematics lead us to this era of “Machine learning”, which is proving its worth as the data is getting bigger and bigger and going out of reach of human’s computational power.

The intent behind the chosen research topic is to get more deep into these demand forecasting methods and find some reasonable ways to do forecasting. It is impossible to focus into every types of demand forecasting as the numbers of fields related to it is increasing day by day. Therefore, for this thesis transportation demand forecasting was selected and an important section of it (Bike Sharing System) was studied. Bike sharing system is an entertaining, enjoyable and also an affordable way of moving around the town. It is therefore, getting a lot of attention as well as subscribers who uses this transportation medium in a daily basis. Like any other service it is also not free from problems and bike rebalancing is in the top.

Hence, this thesis focuses on the bike share data and try to forecast rental demand from it to find a way to solve this bike rebalancing problem. The proposed method is a Convolutional Neural Network based [7] forecasting technique to forecast hourly bike rental demand for both genders (male and female). This forecasting is also described as a regression problem with the time series feature and these features were restructured as images in order to be trained on the CNN model. This method is trained and tested on the historical dataset from the City Bike Trip records of New York City. The evaluation results also shows that proposed method is performing better than the conventional machine learning methods.

1.1.1 Demand Forecasting

Doing demand forecasting or forecasting in general using Big Data, we can achieve reliable estimation by scrutinizing and finding useful hidden patterns in the data. According to [8] data related decision making can improve predictions.

Tucker [9] also thinks that Big Data will be able to predict our every move really soon. Therefore, we end up asking ourselves one question- “what problems are there while trying to forecast with Big Data?” According to [10] the easiest answer is the size, complexity and the speed of data is unmanageable. Also the variations in data and absence of a common structure makes traditional tools and techniques really difficult to use [11]. Which is then poses a big problem for the organizations out there who are trying to forecast different types of demands. For example, organizations like European central bank also organized whole workshop only focusing on forecasting using Big Data.

1.2 Methodologies

Forecasting is everywhere and for many many years people have been doing a lot of different types of forecasting such as - weather prediction, sports outcomes, political and economical event and so on. As we try to predict a lot of different events, therefore, a lot of different methods can also be applied to do that. Using naive instinct, specialist opinions, or using past events data and match with traditional time series and statistical techniques are just a few.

These demand forecasting are continuously improving with the introduction of latest data mining and machine learning techniques but there is no specific method that allows organizations to foresee future ambiguities and risks. Commonly there are two approaches for forecasting demand -

1. Survey Methods
2. Statistical Methods

1.2.1 Survey Methods for Demand Forecasting

To forecast demand for a short term, survey method is one the most common and straight forward method out there. The goal of this method is to get the intentions of the consumers about their future purchase ideas and help the organizations to create better plans. This is done by conducting surveys with consumers about the current product and determine the demand and services to predict the future demand. This method relies on three different exercises such as - Experts opinion poll, Market experiments and Delphi method (a group decision making technique).

1.2.2 Statistical Methods for Demand Forecasting

Statistical methods are complex set of methods which are used to forecast demand in the long run. Demands are mostly forecasted on the availability of the historical and cross sectional data. Unlike the survey methods, statistical methods are reliable and cost effective because of the minimum element of subjectivity. There are a lot of different statistical methods out there like- Trend projection method, Barometric methods and so on. But a very popular and reliable method is to do ‘Machine Learning’ and ‘Deep Learning’ based demand forecasting. A brief details about them is given below.

1.2.2.1 Machine Learning for Demand Forecasting

Nowadays automated detection of meaningful patterns in data is denoted by Machine Learning. The usage of machine learning has grown significantly in the past couple of decades and became a very common tool in order to extract information from a huge sets of data. Machine learning applications are also making our everyday life lot easier, starting from- search engines, which knows how to fetch the best and most relevant results. Anti-virus, spam, fraud softwares- to filter our data, mail,

credit card fraud detection, autonomous or collision prevention on cars, as well as in a lot of applications on the fields of medicine, astronomy etc. In this case machine learning methods has been a great success for demand forecasting based on historical data.

Before going too deep into machine learning we also should know when we need machine learning. We mainly need the shelter of machine learning when a task becomes too complex to program.

- These kind of task may be performed by humans or animals such as driving, speech recognition or image processing. Therefore to deal with this kind of problem, machine learning tools comes in handy, as it can learn from past experience or data in order to provide adequate results.
- These task can be beyond human capabilities- such as analysis of big complex data sets of astronomic data, predicting or forecasting future, converting medical data into knowledge, recommendation system etc.

There are two different types of machine learning tool -

1. Supervised: When the manually labelled dataset is provided to carryout future work.
2. Unsupervised: to describe the hidden structure of the unlabeled data.

In this thesis, I have dealt with supervised machine learning where I used labelled bike sharing service data to do the forecasting task of the system.

1.2.2.2 Deep Learning for Demand Forecasting

Trying to mimic the human brain with the robustness and efficiency and adeptness it exemplifies has always been the main challenge in AI (Artificial intelligence)

research field. Humans are capturing countless data every second and still capable of finding the good use of those in future use with a crisp method. And in neuroscience research they are developing new ideas for designing systems which represents various types of information based on the findings that provides insights of the representation of the human brains. This new method leads to the breakthrough of the Deep learning which emphasizes on the computational models for useful knowledge representation just like the neocortex. In neocortex rather than explicitly processing signals it permits them to proliferate through a very complex order of parts in a way that overtime it will learn to provide observations based on their regularities. Therefore, deep learning can produce a new representation of the data and solve very complex forecasting models like demand forecasting.

In simple words, Deep learning is a complex level of machine learning algorithm where -

- It consists of many levels of layers which are used to extract features from the input. Every layer uses its previous layers output as their input.
- It can be supervised or unsupervised.
- It forms a hierarchical representation by originating higher level features from the lower level features.
- The composition of these layers are dependent of the problem it is trying to solve.
- At each layer, like an artificial neuron the signals are transformed and the parameters of these neurons are learned through the training of the network.

Based on the application and framework it is being used every deep learning algorithm has their own strength and weaknesses. In this thesis I have used one of

the well established and popular method Convolutional neural network (CNN) to solve the forecasting problem of the bike sharing system.

1.2.2.3 Convolutional Neural Network

To solve the goal of the thesis the method I used is the Convolutional neural network which is a very popular deep learning method. It is a kind of feed forward artificial neural network where the neurons connectivity patterns are enthused from the visual neocortex of the animals [12]. This biologically inspired machine learning method is included with some self optimizable neurons through learning. Starting from the input vector to the output result, the whole network states a single weight function by keeping the traditional neural networks regular ideas and adding some special features to it.

Architecture: One of the main difference between the artificial neural network and CNN is, CNN consists of neurons of three dimensions [figure 1.1]. These are the spatial dimensionality of the input width and height, and the depth. Also the neurons of a CNN in any specified layer is associated to a minor area of the layer prior to it.

In overall, CNN is consists of three different types of layer - Convolutional layer, Pooling layer and the Fully connected layer.

The key functionality of the CNN layers can be described in the following areas:

- The input layer of CNN will work as same as the ANN input layer by distributing the input values to the next level of layer.
- The next layer of the CNN is the ‘Convolutional layer’. This layer regulates the output of the neurons connected to the local area of the input. This process is done by the scalar product of the local input area and the weights. Each

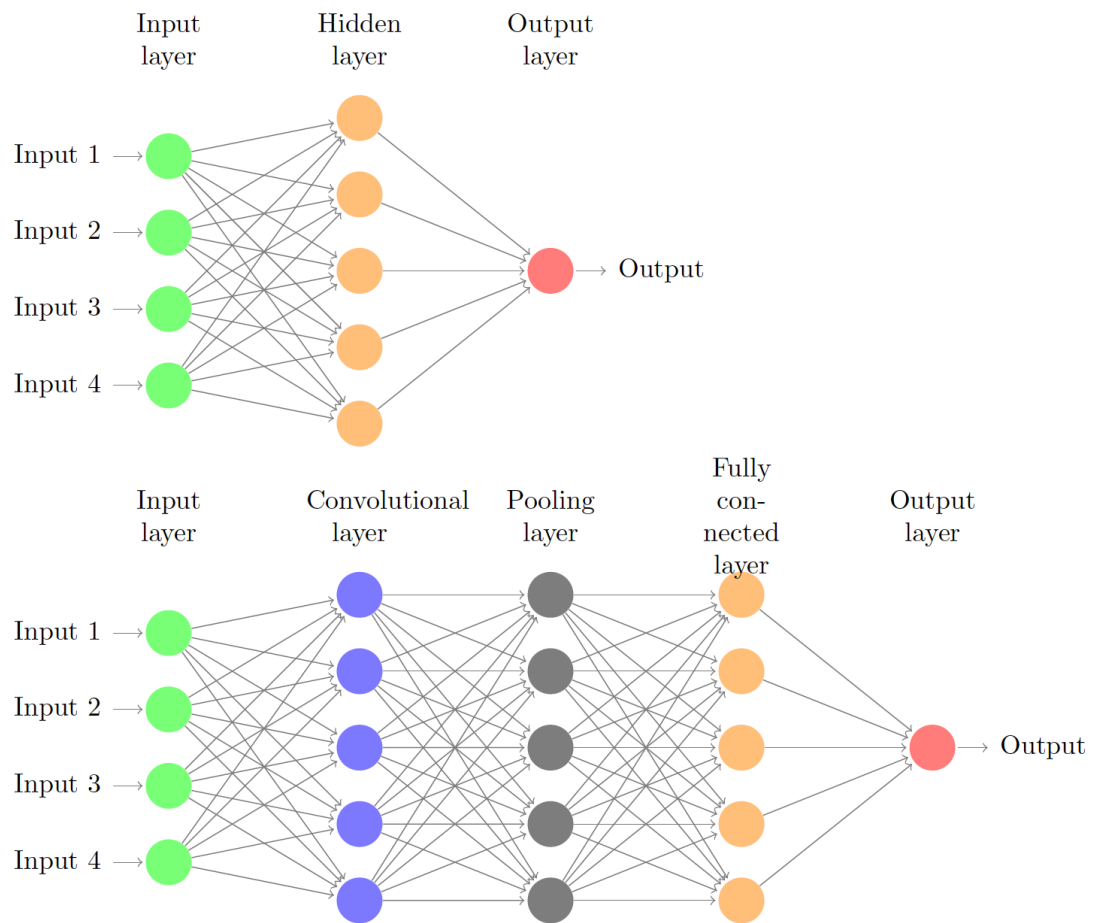


Figure 1.1. Architecture of the traditional Artificial Neural Network (ANN) (above) and the Convolutional Neural Network (CNN) (below).

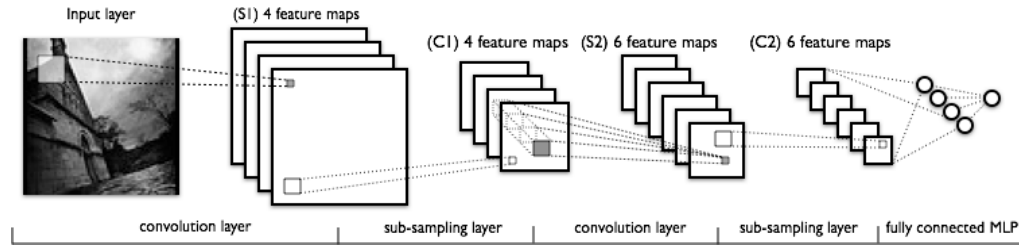


Figure 1.2. A sample Convolutional neural network architecture [1]

neuron of this layer takes input from the previous layers rectangular section. Therefore, convolutional layer is an image convolution of the previous layer.

- The down-sampling by the spatial dimensionality of the connected input area is then performed by the next layer which is the ‘Pooling layer’. This layer takes a small block from the convolutional layer and produces a single output by subsampling. Therefore, this layer reduces the number of bounds are placed in that activation.
- ‘Fully connected layer’ performs just like it works on a traditional artificial neural network. This layer produces the class scores of the activations. This scores are then used for doing different types of classifications. Also in order to improve overall performance ReLU (rectified linear units) can also be used between these layers.

More mathematical operations about CNN are discussed in the later part of the thesis.

1.2.3 Big Data & its Challenges

In order to state a huge amount of data which can not be handled by the traditional data management methods because of the size and the complexity of the data, the term “Big Data” was first introduced by Roger Magoulas from OReilly

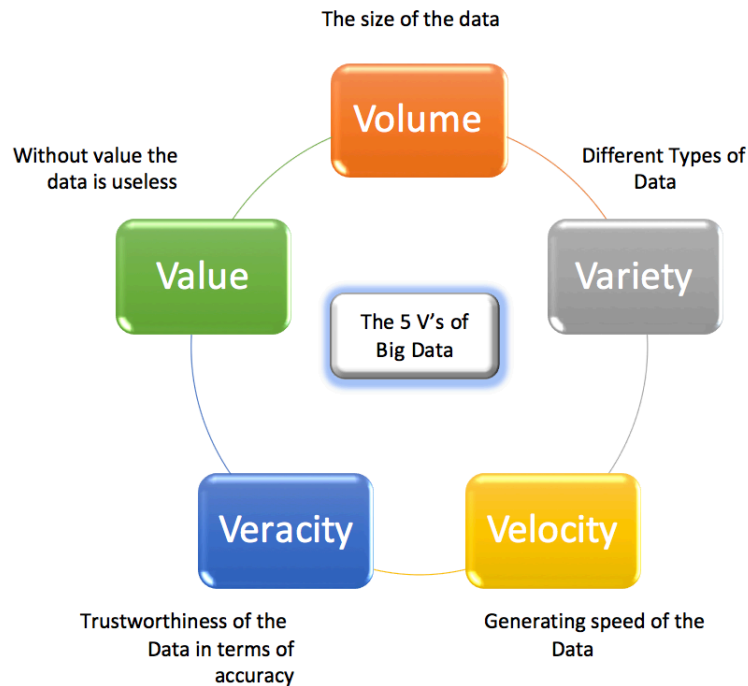


Figure 1.3. 5 V's of Big Data

media in 2005. According to Sam Madden an associate professor of MIT, Big Data is “data that’s too big, too fast, or too hard for existing tools to process.”

To overcome this Big Data and extracting information from it we should discuss the 5 V's of the Big Data. These 5 V's includes describing the size of the data (Volume) of different categories (Variety) below the manageable rate (Velocity) and the authenticity (Veracity) as well as the importance of the data(Value) [figure 1.3].

In order to forecast future, some important challenges often develops from the nature of the data. Few of them are listed below.

1. Statistical impact: Without getting mislead by randomness, it is important to get precise and significant statistical results.
2. Analytics construction: There is no fixed rule or way to decide the optimal

architecture of the analytic system.

3. Disseminated mining: To have distributed versions of the methods used, it is must to carryout a lot of research work and theoretical and practical analysis.
4. Data compression: Storing big data efficiently is also a very important challenge when it comes to Big Data. We can either store the important parts of the data or we can sample the data in a way where we think data is more relevant.
5. Time changing data: Sometimes data changes a lot over the time, for this thesis the data changes significantly over the time. Therefore, it is important for used methods to adapt over the time.
6. Visualization: Another important challenge with dealing with big data is to analyze how to visualize the data to find/solve the problem.

By overcoming these challenges my goal was to forecast the bike share usage of New York City as well as to find different ways to improve this bike share service. In my thesis, I have considered a set of bike share usage data as image processing problem and in the following part of my thesis I will discuss more about my method and why I chose this approach.

1.3 Thesis Outline

This thesis is structured in the following sections :

Chapter 2, begins with a review of demand forecasting with Big Data and the problems and potentials are presented, as well as some challenges and applications. This chapter also provides some reviews on the approached topic of bike share service demand forecasting. Chapter 3, explains the problem of bike share service and then presents the details about the dataset which was used to build the forecasting model.

Furthermore, a few preliminary analysis of the data as well as the methodology of the proposed model is described. Chapter 4, presents the feature selection process of the data and brief explanations of the baseline models and evaluation metrics followed by the results and discussions of the proposed model. Chapter 5, summarizes the contribution of the thesis as well as some future works are mentioned.

CHAPTER 2

LITERATURE REVIEW

2.1 Theory

As previously mentioned, forecasting using Big Data is a ground-breaking sensation which is one of the most discussed and one of the most important topic in the modern age. In this section a detailed review of the use of Big Data for forecasting with identifying the problem, potential & challenges come with it also the related applications is presented. For obtaining meaningful forecasting results from Big Data is often faced with a lot of challenges such as- skills used, software and hardware, architecture of the algorithm, signal to noise ratio, statistical importance as well as the nature of the data itself. Even though all this challenges are continuously trying to impede the forecasting, many fields like Energy, Economics, Population dynamics are the main exploiters of it. With some common tools like Bayesian, Regression, Neural Network models are implemented for these type of forecasting with the availability of Big Data.

2.1.1 The Basics

The arrival of Big Data is becoming history now as industries around the world are generating enormous amounts of data every second. Therefore it is very important for organizations to respond to this digital growth of asset and adopt necessary tools to exploit it for their own benefit. As [13] states that there is a need for adoption of data mining techniques which can support in modelling the complicated connections of the characteristics present in Big Data. Also in current years, organization felt the importance of the risk management due to the recent financial crisis.

According to [14] corporations are now trying to minimize associated threats of maximizing their opportunity by using risk management tools. Hence, forecasting using Big Data has the potential and the ability to boost the organizational performance whereas empowering enhanced risk management [15]. Conversely, Not every author is agreeing with the Big Data sensation. In [16] Poynter describes that Big Data will be more intuitive for connecting dots rather than painting a whole new picture. According to [17] the year 2013 was the year to get familiarized with Big Data and 2014 is the year to start exploiting the Big Data for profitable gains. This chapter will provide a brief knowledge about Demand Forecasting in current time using real world Big Data also the challenges and applications of it.

2.2 “Potential” and “Problem”

Demand forecasting using Big data has a lot of different opportunities for lucrative gain. Companies have already started exploiting Big Data in order to forecast demand and work accordingly in order to improve better management and also increase revenues. Fashion companies like EDITD are trying to forecast future fashion trend demands by using the data collected from social media. Netflix used Big Data for decision making before starting their own TV show House of Cards which then became very successful. Airline companies are also relies on the demand forecasting very much. Therefore, we can clearly see that the potential behind demand forecasting using Big Data are huge and actually overwhelming. But this is so astonishing that sometimes it becomes scary as in [18] author described a story where a woman complained about a Target store for sending her high school daughter pregnancy related coupons. But a few weeks later she apologized as her daughter was in fact pregnant [18].

2.3 Challenges

Demand forecasting using Big Data poses a lot of challenges which we must overcome in order to get any meaningful outcomes from the data. It is also vital to know that Big Data's obtainability does not end the problems [19], therefore, in this section I will be primarily focusing on the challenges. Some of the challenges are associated with the techniques, assumptions [20] while [21] suggests that absence of theoretical support of Big Data is also a major concern.

2.3.1 Skills

One of the major challenges while trying to forecast demand using Big Data is the expertise needed and the availability of experts in this specific task. [11] shows that the importance of cutting-edge skills in order to deal with Big Data and [22] states the shortage of the data scientists who are able to tackle this problem. Traditional researchers and scientists are working on old methods for decades and now this emerging era of Big Data is a challenge in itself. Therefore, inappropriateness of these traditional methods which were used on the old data types are hampering current world from forecasting using Big Data [23] [11]. For more than 50 years these traditional statisticians are working on those outdated techniques, which makes it very difficult to develop new skills for Big Data forecasting [24]. Hence, it is necessary for current big educational organizations all over the world to modify their curriculum to adopt the need for new generations of statisticians by include the skills needed for understanding, analyzing and forecasting with Big Data. Thus we can make sure that in future we will have better skills to tackle this new very powerful asset of Big Data and also will be able to forecast more important insights from the data.

2.3.2 Unwanted Data / Noise

Unwanted data or noise is actually making it very difficult to forecast something accurately using Big Data. In [20] the author points out at this problem saying an increasing noise to signal ration is present in the current Big Data. As the data gets larger and larger, exploiting this data is also getting more difficult. Before the era of Big Data, traditional methods used to work relatively well in those smaller sets of data set, but now the Big Data sets consists of a lot of unwanted data or wrong data which makes forecasting very complex. Therefore, New techniques and methods needs to be applied in order to filter out those noises and restructure the data in a way so that we can get the most out of it. There are few techniques already present like- singular spectrum analysis which tries to filter out the noises from a time series and reform the time series with much less noise, which is also proved in [25] [26].

2.3.3 Hardware and Software

Almost all of us who is involved with statistical analysis have faced the problem that programs crashes after doing few thousands to iteration. Like [11] says existing software to do statistical analysis are not powerful enough to do Big Data forecasting where [27] says supercomputers are needed. Capable softwares are also as important as hardwares and nowadays a lot of forecasting techniques are developed and some of them can automatically forecast problems [28]. But these programs are not as reliable when it comes to deal with Big Data. Therefore, to make a powerful forecasting software it needs the required computing capability and only then we can effectively deal with forecasting as the data is getting bigger and bigger.

2.3.4 *Algorithm Architecture*

Current data mining methods are often used for smaller datasets and are unable to perform in a satisfactory manner. Also many time this technique can not deal with the data which are not present in their main memory so data movement between locations are often poses a big problem [29]. Parallel programming or lambda architecture mentioned in [30] are samples of current ongoing research happening on this topic.

2.3.5 *Statistical Importance*

While doing demand forecasting or forecasting with Big Data, it is very easy to make incorrect findings. The amount and the randomness of the data and the noise available in it increases the chance of getting a misleading forecasting report. On top of that finding an appropriate technique has always been a very difficult job for big data. In [31] author also talks about the threat of making false discoveries.

Apart from all these the phrase Big Data itself is a very difficult challenge while doing demand forecasting, due to its in-built characteristics. Very complex and unstable structure of the data makes it tough for data scientists to get any forecasting which is free from poor out of sample outcome [24]. Also it is very difficult to chose the correct method and while using deep learning it is very difficult to chose the number of hidden layers it should have due to the shapelessness of the data as it is continuously changing in real time.

2.4 **Applications of Demand Forecasting**

In this segment, we introduce the present applications which uses demand forecasting by exploiting data mining techniques on Big Data. The main reasons behind doing demand forecasting are :

- Cost reduction
- Agility in the market
- Estimating the financial capability
- Planning ahead to meet necessities
- Improve accuracy
- Visualization - Relationships between different factors

2.4.1 Energy Demand Forecasting

Smart meters and advanced technology in the field of energy are the reason behind the huge data sets that are available currently. These Big Data have opened the door towards endless opportunity for improvement in better energy management and conservation system as well as new energy systems. A lot of experiments have been carried out in the field of energy like, in [32] Nguyen and Nabney tried to exploit the Big Data from the British Energy organizations and they used few different models to evaluate the wavelet transform like, linear regression, GARCH, MLP and Radial basis functions. They found that MLP or GARCH with wavelet transform are the best models for forecasting electricity demands as well as the gas price.

2.4.2 Health Care Demand Forecasting

By using the Big Data in health care field it is possible to improve charge capture and also can reduce the denial rate of reimbursement. According to [33] it is also possible to do location based forecasting in order to coordinate surgeon preference information and also forecasting demand based on the patient demographics, clinic scheduling as well as seasonal demands. These results provide lower cost for the

preparations of cases, reduces inventory levels in OR, supports clinics for much better planning and budgeting to introduce new or to expand current surgical agendas also improving the service and fill rates which leads those clinics to improve patient satisfaction, improved rating and productivity as well as patient safety.

2.4.3 Transportation Demand Forecasting

Demand forecasting in transportation is the method to predict the the number of people or the transportation mediums will be used in the future. This forecasting methods can then be applied on a range of different situations from forecasting traffic volumes on a particular road on a given day to forecasting passenger counts on trains or to forecasting number of ships will be there in a port. The reason behind these forecast is to predict what will be the demand in the future might be and also to specify the standard for planning and design to the most efficient transportation system. Application of demand forecasting in transportation allows us to improve overall transportation system, planning ahead to meet the requirements, developing the infrastructure capacity and design estimation. These can then make sure the approximation of the financial capability of the system as well as the environmental effects it has on them.

In this thesis, I have used the Bike sharing system data of the New York city in order to forecast the demand of bike sharing service and to find ways to improve the service so that clients can get better experience from it.

2.5 Transportation Demand Forecasting : Bike Share Service Approach

Forecasting ride sharing system usage is becoming very popular with the growing of industries like Uber-lyft. But substantial traffic in busy cities and aspiration of an environmental friendly way of transportation made bike share system very popular

and attractive in recent years. Thus Bike share systems provide users an alternate and ecological way of conveyance by dodging personal car usage or taxis and staying stuck at traffic for a long period of time. This growing demand of bike sharing system making big cities to open their own bike share service for which currently more than 600 cities all over the world has the bike share program, allowing their residents to go from one place to another by renting bikes. Jensen et al. [34] also showed (in his case for Lyon) that sometimes it is faster than using personal car to go to places. Except the expenditure almost all the bike sharing program shares the similar technologies and it will be really exciting to predict the future usage of this amazing program and thus allowing users to plan their everyday trip more accurately with satisfaction.

In May 2013, Citibank started a bike sharing service in New York City which is called City Bike. This is the nations leading bike share service with more than 10,000 bikes with 600 stations across the city where customers can rent a bike from any of those station for a quick, fun and reasonable ride on an as-needed basis. Users can obtain membership to rent the bikes and return the bike in an automated process via a network of stations throughout the city. City Bike share recently hit 10 million trips in a year and thus the system generates a lot of data which is very eye-catching field for researchers as it explicitly contains data about the departure station locations, age group and gender of the users, duration, arrival location and so on. Therefore, this service works like sensor network and can be used to study the movement flexibility in a particular city.

Researchers around the world did some related research on the bike sharing system where they all tried to predict that a user can find a bike whenever he goes to the station.

2.5.1 *Related Works*

Bike sharing system first started in Amsterdam in 1965 with the 1st generation ‘White bikes’ and since then the popularity is increasing all over the world [35]. Due to technological advancement the 3rd generation bike sharing system is available now, where it is possible to do bicycle reservations, information tracking, automated pickup and dropoff system [36]. This IT based integrated system for bike share programs made it possible to collect its data to do research and find possible ways to improve the system.

Researchers around the world did some related research on the bike sharing system and these are mainly focused into four different problems. Strategic program scheme, quality improvement, demand analysis and rebalancing issue. Almost all the major cities around the world already planned or is planning or considering execution of this bike share system. For the first problem, estimating potential demand in order to optimize location, deellolio et al. [37] presented an inclusive approach for implementation. Location is also another issue which was widely discussed in Martinez et al. [38] and Prem Kumar et al. [39] where they propose MIP models to tackle this issue. Also a trade off model for the interest of both customers and sponsors were proposed in [40].

There has been also several research focusing on the second problem which is the quality improvement. Raviv et al. [41] proposes a model for probable users displeasure in every station based on a queuing system with finite capacity. Nair et al. [42] and Nair and Miller [43] decompose system-wide reliability into a set of dual-bounded chance constraints for every station. By assumption of waiting customers [44] shows bike sharing stations as a dual Markovian waiting system.

Similar to this thesis there has been few research about Demand analysis where

the analysis is mainly divided into two sections 1) forecasting future demand, 2) understanding factors for administrative decisions. Kaltenbrunner [45] tries to predict the system inventory state also suggests making this information available to users. Rixey et al. [46] investigates the effects of demographic and built environment characteristics near three bike sharing stations on bike sharing ridership levels in three operational U.S. systems. The regression analysis identifies a number of variables as having statistically significant correlations with station-level bike sharing ridership and identify station locations that will serve the greatest number of riders. Borgnat and Abry [47], Froehlich and Oliver [48], Lathia et al. [49] and Borgnat et al. [50] identify a temporal demand pattern and forecast the number of rentals. Giot and cherrier [51] attempts to predict the bike share usage up to one day ahead with an hour of granularity and compared how adaboost regressor, ridge regression, SVR, random forest nad gradient boosting regressor performs on dataset on two years. Their results sows that most regressors are sensitive to over-fitting.

For the fourth problem, Shu et al. [52] showed that estimate rebalancing operations can lead to an additional 15-20% of trips supported system-wide. Chemla et al. [53] present a branch-and-cut algorithm for the single-vehicle problem, with results on instances of up to 100 stations. Lin and Chou [54], rather than simply using Euclidean distance they proposed an optimization way taking road conditions, traffic regulations, and geographical factors into account. Their actual distance path calculation is used to implement a exploratory for the VRP. Naturally, using actual distances would lead to decreased costs in practice. Yoon et al. [55] proposes auto regressive moving average model to help bike share users to navigate around the city. By giving the starting and end point the application can suggest which stations would be best to use by considering the walking time and biking time and available return slots at the end station. This model is useful for stationary signals and the

bike share system which is varying over time.

From the researches mentioned above it can be observed that there are many research going on in different topics related to bike share system but for the forecasting bike usage people are mainly using machine learning models with space and time series features and building applications to help customers navigate around the city easily. For this thesis, I present a Convolutional Neural Network based demand forecasting model where the bike usage data is treated as images and solved a regression problem using CNN. Results also shows the efficiency of the model.

CHAPTER 3

METHODOLOGY

3.1 The Problem

Like any other program this bike sharing service also suffers few major problems. Bike rebalancing problem being the most important and significant of them all. This rebalancing problem occurs when a bike station is either full or empty and makes it difficult for users to rent or to park their rented bike. Thanks to the current prediction systems, it is possible to predict the usage of the bike share program and therefore allowing the users to have better experience of the bike share system. The objective is to make sure the users can walk to a bike docking station and find bikes to rent and also to discharge the rented bikes without having to wait for another user or going to another station to do that.

In order to find the problem and to set the goal of my thesis, first I did some statistical analysis on the bike share data and from those studies I saw that the distribution of the data is not balanced and same goes for gender distribution of the subscribers. Therefore, I tried to find a solution to overcome this two problem which is to forecast hourly rental bike demand for both men and women. I treated this problem as a regression forecasting with time series features and proposed Convolutional Neural Network (CNN) to do the forecasting which outperforms other traditional machine learning methods. I have used the data obtained from the City Bike system and weather data to forecast hourly rental demand of the Bike share system.

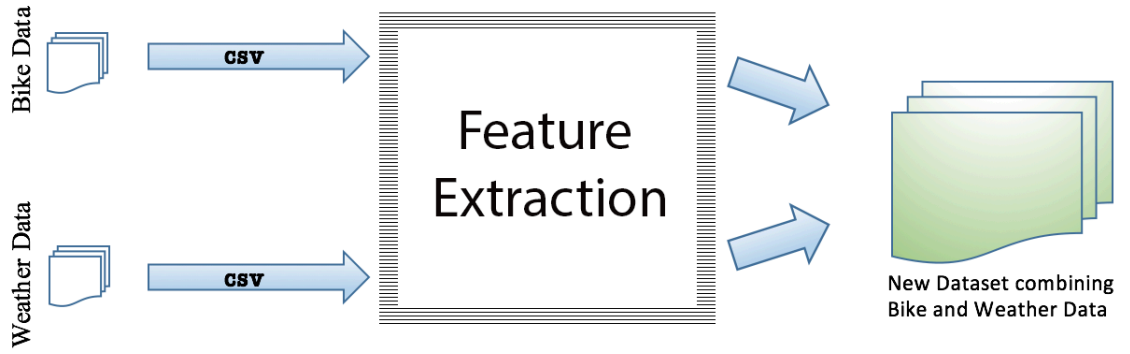


Figure 3.1. Dataset Overview.

3.2 Datasets

The goal of this thesis is to forecast bike demand using the users historical bike usage patterns along with the weather data and user's information in a time series arrangement. Therefore, this thesis attempts to answer the following questions : “How many bikes are needed every hour?”, “Who are renting those bikes?”, “How weather is affecting the bike share usage?”, “Is it a holiday, weekend or normal weekday?”. Therefore, datasets should contain three major information: 1) bike renting time and details, 2) users information, 3) weather data as well as 4) holiday/weekday-weekend information.

Collected dataset on the city bike share system contains details about the trip and the user who is renting it from January 2015 to July 2015. I have also collected weather datasets during that period and used January 2015 to June 2015 data sets to create the training set and the rest for testing set.

Figure 3.1 shows the overview of the dataset preparation to build the forecasting model. Detailed information about the dataset extracting and processing will be discussed later in this chapter.

Table 3.1. Details of the Citi Bike Trip data

Field name	Type	Explanation	Information
tripduration	Numeric	Duration of the ride in seconds	1
starttime	Date	Start date and time of the trip	1
stoptime	Date	Stop date and time of the trip	1
start station id	Numeric	Starting station id	1
start station name	Character	Starting station name	1
start station latitude	Numeric	Starting station latitude	1
start station longitude	Numeric	Starting station longitude	1
end station id	Numeric	End station id	1
end station name	Character	End station name	1
end station latitude	Numeric	End station latitude	1
end station longitude	Numeric	End station longitude	1
bikeid	Numeric	Bike identification	1
usertype	Character	Type of user (Subscriber = Annual member, Customer = 24 hr/ 7 day pass)	2
birthyear	Date	Users birth year	2
gender	Numeric	Gender of the user (0=unknown, 1= male, 2= female)	2

3.2.1 New York City Bike Dataset

The city bike system website provides historical datasets for all the researcher interested to carry out their research. This dataset is in CSV - comma separated values format. This dataset contains 6 key variables which are key to this thesis - ‘tripduration’, ‘starttime’, ‘stoptime’, ‘usertype’, ‘birthyear’, ‘gender’. This deals with the first two major topics which mentioned are above 1) Bike renting time and details, 2) Users information.

The City Bike dataset has 3,379,904 entries of bike rental data and it all has the content showed in table 3.1. Target bike rental dataset is then designed to get the time information, which is found by splitting the ‘starttime’ into ‘hour’, ‘month’,

Table 3.2. Target dataset of the Citi Bike trip data

Source	Name	Type	Information
starttime	hourly count	numeric	Total number of rented bike per hour
	hour	numeric	Hour data
	day	numeric	Day
	month	numeric	Month
usertype	subscriber	binary	Subscriber count
	customer	binary	Customer count
birthyear	age	numeric	Age of the user
gender	unknown	binary	Unknown gender information of the user
	male	binary	Male count
	female	binary	Female count

‘day’. Total number of rented bikes per hour can be then grouped in one field. Table 3.2 shows the extracted details from those bike datasets which is then used for the experiment.

3.2.2 New York City Weather Dataset

For this thesis I have collected some historical New York City weather data from the NOAA (National centers for environmental information) website. From this dataset 8 key factors were chosen. These are Date (Year, Month, Day), snow depth, snow fall, average wind speed, trip, precipitation. The target dataset from the weather data is shown in Table 3.3.

3.2.3 New York City Holiday Dataset

I have also collected dataset of the holiday during the period of Jan’2015 to July’2015 shown in figure 3.2 and used it to build the training set of my model. There were 8 national holidays during the time period of the datasets related to this experiment of the year 2015 in USA. With this holiday weekend and weekday binary values will also be included in order to create the target dataset which is shown in

Table 3.3. Target weather dataset

Name	Type	Information
Year	numeric	Year count
Month	numeric	Month count
Day	numeric	Day count
Snow depth	numeric	Snow depth level (in)
Snow fall	numeric	Snow fall rate (in)
Average wind speed	numeric	Wind speed (mi)
Temperature	numeric	Temperature (F)
Precipitation	numeric	Rain, snow, sleet, or hail that falls to the ground (in)

Table 3.4. Target Holiday dataset

Name	Type	Information
Holiday	binary	Holiday binary value
Weekday	binary	Weekday binary value
Weekend (not holiday)	binary	Weekend which are not holiday binary value

Table 3.4.

3.3 Statistical Analysis

In order to find the problem and to set the goal of my thesis, first I did some statistical analysis on the bike share data and from those studies I saw that the distribution of the data is not balanced and same for the distribution of the gender of the subscribers. Therefore, I tried to find a solution to overcome this two problem which is to forecast hourly rental bike demand for both men and women. I treated this problem as a regression forecasting with time series features and proposed Convolutional Neural Network (CNN) to do the forecasting which outperforms other traditional machine learning methods. I have used the data obtained from the City Bike system and weather data to forecast hourly rental demand of the Bike share

 **Federal and state holidays of New York, USA in 2015**

Day	Date	Holiday	Comments
Thursday	January 01	New Years Day	
Monday	January 19	Martin Luther King Day	Third Monday in January
Thursday	February 12	Lincoln's Birthday	Connecticut, Illinois, Missouri, New Jersey, New York.
Monday	February 16	Presidents' Day	3rd Monday in February. Not all states
Sunday	May 10	Mother's Day	2nd Sunday in May. Not a public holiday
Monday	May 25	Memorial Day	Last Monday in May
Sunday	June 21	Father's Day	3rd Sunday in June. Not a public holiday
Friday	July 03	Independence Day (observed)	Friday is a holiday as 4th July is on a Saturday

Figure 3.2. Holiday list of NYC in 2015 (Jan-July) [2]

system.

Getting more insights and understanding of the data before applying any machine learning tools is very important when it comes to forecasting. First we saw that there are two types of users in those data - subscribers and customers. So, then I tried to plot the trend of the bike share rentals between them.

From this figure 3.3 we can see that the customer count is almost less than 1/3 in count and the distribution of the data is unbalanced which can be because of a lot factors like weekend, weekdays, good or bad weather, holiday etc.

3.3.1 Weekly Distribution

To get more deep into this unbalanced distribution of the data, more deep analysis of the data is needed. Rather than getting the rental demand trend in

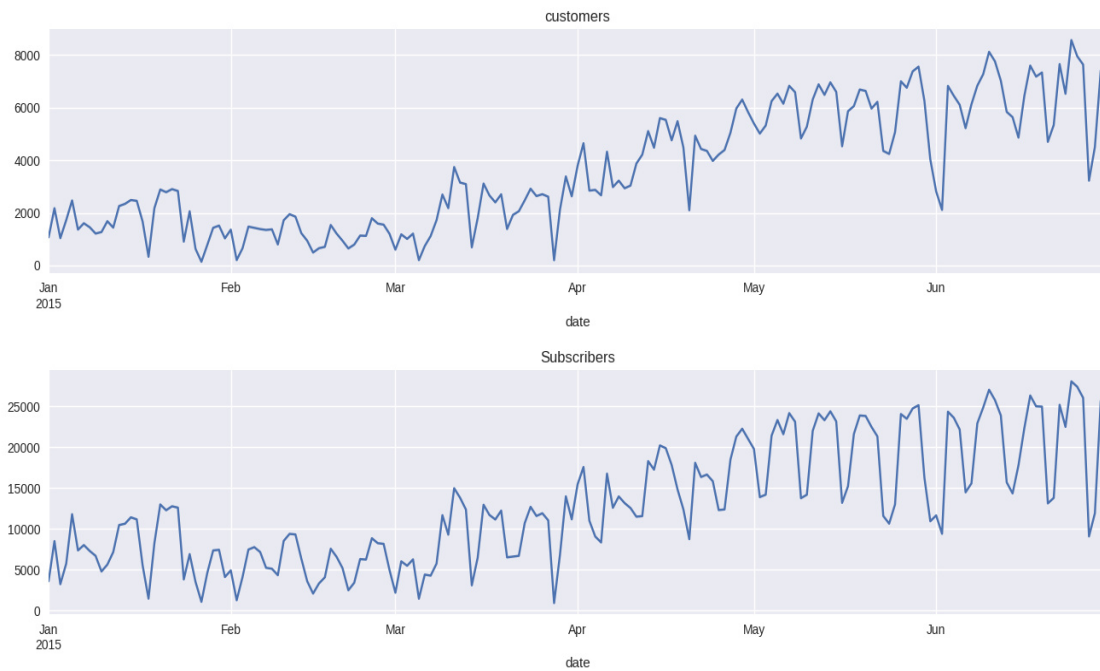


Figure 3.3. Customer vs Subscribers rental trend

monthly basis, weekly distribution is then applied. This time the weekly data is divided in gender basis, which is another important factor for this thesis. figure 3.4 shows the weekly average demand trend and from this figure we can see that rental demand significantly decreases when the weekend comes. This figure also shows a very valuable comparison between “male” and “female” rental demands. Females are 1/4 less likely to rent a bike than males in New York City. There are also some unknown data which are the users who did not provide their gender details. The unknown data was really low during the weekdays but in the weekends the count rises high enough to even surpass female counts. Which also tells us that during weekends alot of people who are not subscribers, likes to roam around in the city using this bike sharing system.

Doing daily analysis it was also found that the average number of bike rental

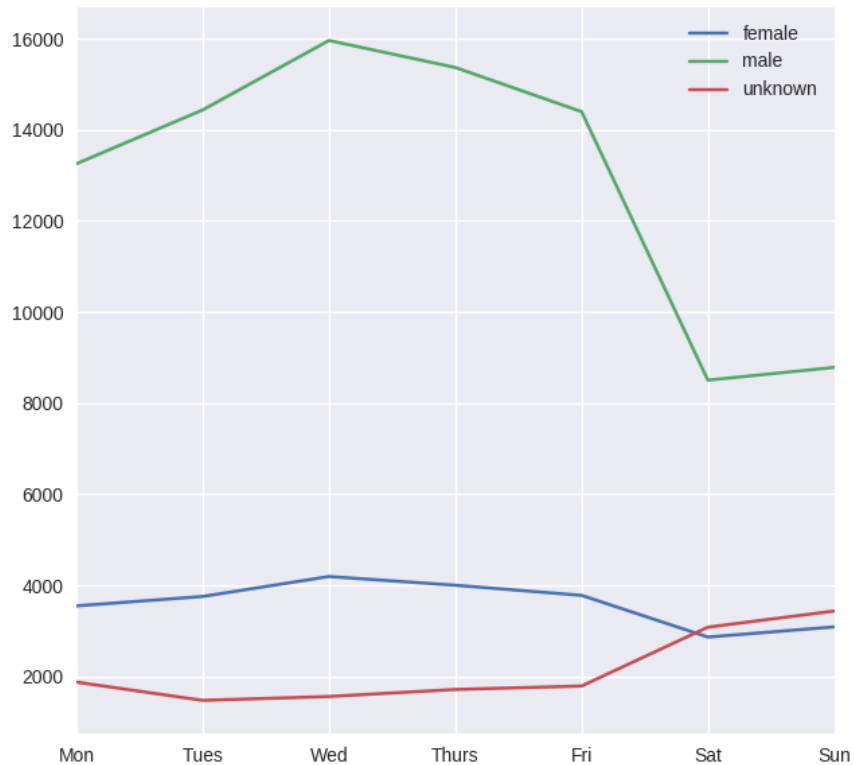


Figure 3.4. Male vs Female average rental demand per week

demand per day for male is about 13 thousands where for female it was 3,602 and for unknown data it was around 2 thousands. Another gender distribution is also shown in the figure 3.5 where 69.3% of rental data is provided by male and only 19.3% of them are female, rest 11.4% being the unknown categories.

3.3.2 Trip Duration Frequency

A detailed analysis on the trip duration is also carried out in order to understand user behaviors behind renting bikes from this bike sharing system. There is also an extra charge after 30 minutes of use of bike in one go for customers and after 45 minutes for subscribers as shown in figure 3.6 . So therefore, users are more likely

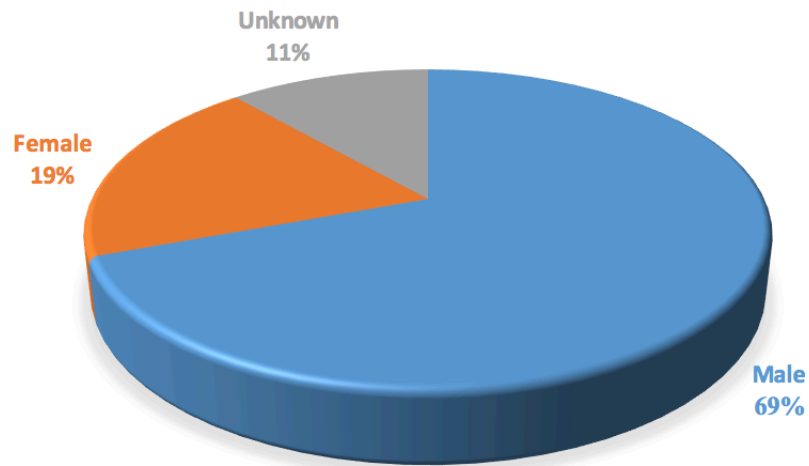


Figure 3.5. Gender distribution in rental demand of bikes

to avoid this extra charge as one of the main reason behind using bike share system is to move around the city in an affordable way. But, from the data we observed an interesting tendency of the users. The maximum duration being 97744.35 minutes which is either lost and then found or just an outlier in the data which we can neglect and minimum duration found is a minute. The mean duration is around 12.5 minutes and most importantly more than 75% is around 14 minutes. So, from this we can realize that people mostly uses bike share program for short trips in the city.

3.3.3 Hourly Distribution

Comparing the plots of the figure 3.7 and figure 3.8, the shown graphs are based on the details about customers and subscribers and it shows that the pattern of the plot is noticeably different. “Subscribers” who holds the annual pass for the bike share system displays ‘working’ pattern. Which means the bike rental demand during the start and the end of the working hours are huge. Most people rents bike

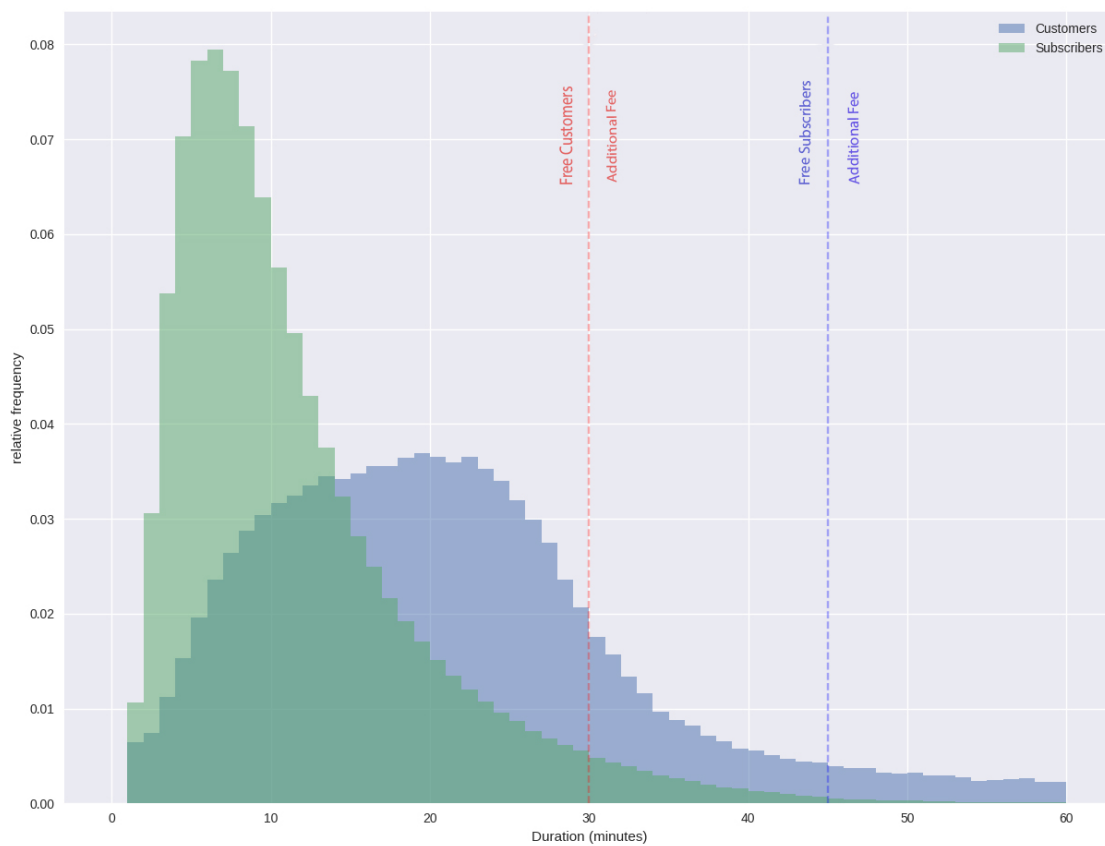


Figure 3.6. Trip duration distribution between subscribers and customers

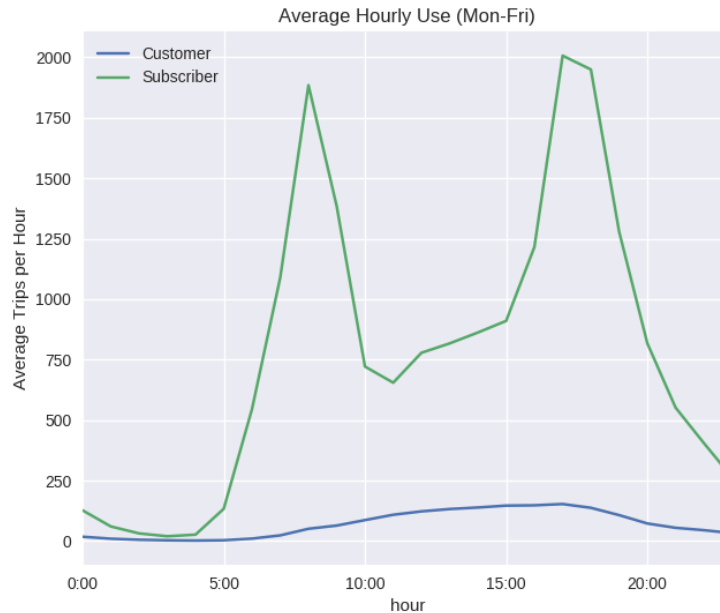


Figure 3.7. Hourly average rental demand in the City Bike trip data (Monday-Friday of Jan-June'2015)

around 8:00 am and from 4:30 to 5:30 pm. “Customers” who uses day or weekly passes shows random distribution of the data. It is also found that the maximum hourly demand of bikes in NYC in those time period is 3034 bikes and averages in about 540 bikes per hour, minimum being 1 bike.

It is also found that the maximum age of a renter is a male being 77 years old and minimum is a female of 17 years old. The reason behind the minimum year is 17 is maybe because people under 17 do not hold credit cards. It is also very pleasing to see the mean values of age in general is 39 years.

After all this analysis of the data we can then conclude as the problem, where in these short bike trips between places it is very important to make sure users can rent and return a bike whenever they want. Therefore, the attempt of forecasting bike rental demand can play a vital role in order to predict the demand before it

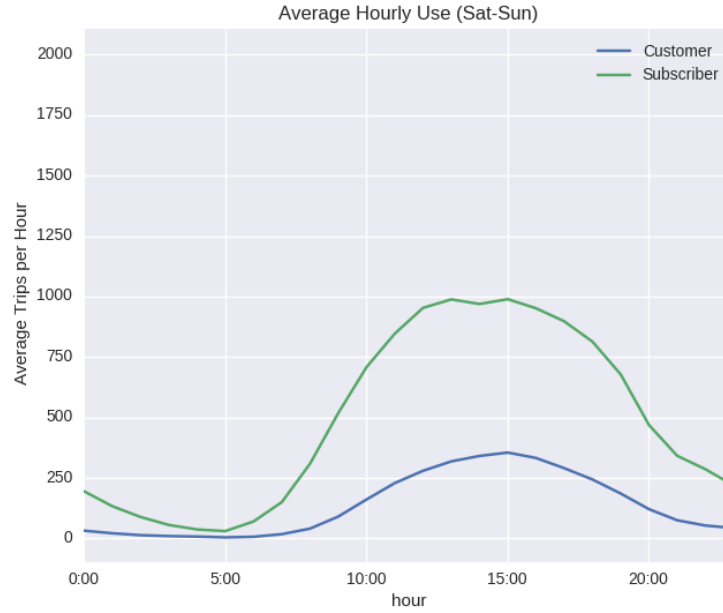


Figure 3.8. Hourly average rental demand in the City Bike trip data (Saturday-Sunday of Jan-June'2015)

happens and thus rebalancing the whole system efficiently. Data also showed that customers normally use bikes for short trips. So failing to deliver a rentable bike or a parking place in a desired station can cause the system to fail as a whole, as customers can sometimes just walk to the destination. Hence, it is very vital for a bike sharing program to make sure customers can get the best possible experience from it and stay subscribed.

3.4 Methodology

Statistical results on the bike data prove that the distribution of the data is sometimes similar and therefore there is a big chance for researchers to find ways to improve the system. From the figure 3.4, it shows the huge difference in rental demand count between male and female users in NYC. Therefore, there is big potential to increase the number of female users into this bike share system. In order

to solve this matter, different bikes for different genders can be a very good option. By making the female bikes more appealing and colorful and also to design the bike to support female's body structure can provide better experiences for female users.

This leads our main problem into two different problems. Forecasting and then providing enough bikes for both male and female users at any given hour. And the amount of female bikes needed can then also be determined from the rental demands of the female users. By doing this we can then forecast the adequate number of bikes needed also that every hour there will be enough bikes available for the users precisely.

3.4.1 Convolutional Neural Network

Before discussing the problem and a way to solve it, it is necessary to understand how the Convolutional neural network works. As mentioned earlier in chapter 1 a CNN has three types of layers and the layer in use will decide the forward or back propagation is needed. A deep explanation of Forward and Back propagation of Convolutional layer and pooling layers are discussed below.

3.4.1.1 Forward Propagation

Here it shows the computation of the layers in case of forward propagation :

Convolutional Layers: To explain this section suppose we have $N \times N$ neuron layer trailed by the Convolutional layer. Then, the convolutional layer output w will be $(N - m + 1) \times (N - m + 1)$ if we use $m \times m$ filter.

So, the pre-nonlinearity input to some unit X_{ij}^l in the layer,

$x_{ij}^l \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} y_{(i+a)(j+b)}^{l-1}$, this is done by summing up the weights by the

filter sections from the earlier layer cells. After that the convolutional layer adds its non linearity $y_{ij}^l = \sigma x_{ij}^l$.

Max - Pooling Layers: Max pooling layers basically takes some $k \times k$ region and outputs the maximum single value for that region. For example, if they are getting input from a $N \times N$ layer then they will output a $\frac{N}{k} \times \frac{N}{k}$ layer and using max function each $k \times k$ block reduced to just a single value.

3.4.1.2 Back Propagation

The computation of the layers for back propagation :

Convolutional Layers: Suppose the error function E and the error values at the convolutional layer is known. The error we know and the calculation needed for the previous layer, with respect to each neuron output is the partial of E with respect to each neuron output ($\frac{\partial E}{\partial y_{ij}^l}$). By applying chain rule we can also compute the gradient.

$\frac{\partial E}{\partial w_{ab}} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial x_{ij}^l} \frac{\partial x_{ij}^l}{\partial w_{ab}} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial x_{ij}^l} y_{(i+a)(j+b)}^{l-1}$. Here, we should add all x_{ij}^l where w_{ab} happens. We also need to calculate the deltas which the values of $\frac{\partial E}{\partial x_{ij}^l}$ by using chain rule again .

$\frac{\partial E}{\partial x_{ij}^l} = \frac{\partial E}{\partial y_{ij}^l} \frac{\partial y_{ij}^l}{\partial x_{ij}^l} = \frac{\partial E}{\partial y_{ij}^l} \frac{\partial}{\partial x_{ij}^l} (\sigma(x_{ij}^l)) = \frac{\partial E}{\partial y_{ij}^l} (\sigma'(x_{ij}^l))$. Therefore, deltas of the current layer can be very easily computed by using the derivative of the activation function ($\sigma'(x)$). Now the gradient can be computed with respect to the weights used by the convolutional layer.

In addition, we need to propagate the errors back to its previous layer and by using chain rule once again we can compute that

$$\frac{\partial E}{\partial y_{ij}^{l-1}} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \frac{\partial E}{\partial x_{(i-a)(j-b)}^l} \frac{\partial x_{(i-a)(j-b)}^l}{\partial y_{ij}^{l-1}} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \frac{\partial E}{\partial x_{(i-a)(j-b)}^l} w_{ab}$$

From this we can see the convolution effect. The equation above is valid for points which are atleast m away from the left and top edge. And to fix that we simply need to fix that with zeros and it will be simply a convolution using w which is flipped by both the axes.

Max - Pooling Layers: The backpropagated error for the max pooling layer is sparse. As mentioned earlier pooling layer do not do any learning themselves. And in forward propagation blocks $k \times k$ is reduced to its single value. Then this value attains error calculated from back propagation from the previous layer. Therefore, the value just gets forwarded to the place where it came from.

3.4.2 Problem Formulation

In this section, previously introduced hourly bike rental demand forecasting will be studied more using the datasets and based on the analysis mentioned earlier. Hourly bike rental demand forecasting has a goal to predict the bike rental demands based on the historical rental demand and the proposed method is a regression problem with the time series features.

Time series is an arrangement of measurements taken at controlled points in time (usually equally spaced manner). It was originally developed by applying statistical methods in the field of econometrics, which is then used in a variety of fields like physics, ecology and engineering. The main focus of the time series analysis is prediction by producing precise forecasting of forthcoming measurements from a given series. Forecasting using time series takes the form-

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

where, X_1, X_2, \dots, X_n are the independent variables, b_0 is the intercept and b_1, b_2, \dots, b_n are the coefficients represents the contribution of X_1, X_2, \dots, X_n .

As we are treating our problem as a time series regression problem, so to do forecasting we need to use related historical rental demand data. For the target value Y which is the bike rental demand, we then have to design a model for Y_t which is the value of rental demand Y at any given time t . Before building the model architecture we need to realize that this time series are related to its previous

observations. Therefore, to build a model for y_t we need to take all the previous values of y_t, \dots, y_n as it has an effect on it. So our model should look like

$$Y_t = df_{model}(Y_{t-1}, Y_{t-2}, \dots, Y_{t-n})$$

Here, df_{model} is the demand forecasting model which is dependent on the previous Y_t values such as Y_{t-1} means last hours rental demand, Y_{t-2} means the rental demand of two hours ago from t hour and so on.

But this df_{model} also depends on the other two datasets mentioned earlier, weather and holiday. So, after taking those in account we can rewrite our equation into something like this :

$$Y_t = df_{model}(W_{t-1}, W_{t-2}, \dots, W_{t-n})$$

Where, W_{t-1} is the combination of all the features that came to consideration while forecasting bike rental demand Y at time $t-1$, such as rental demand, weather, holiday.

After selecting this features and modeling the problem, restructuring was done to treat this features as image pixel features. The reason is to treat this bike rental demand forecasting problem as an image evaluation problem figure 3.9.

In an image analysis problem, image samples contains many many pixels and these pixels are surrounded by a lot of pixels that has influence on the target pixel. For a 3×3 example as shown in the left sub figure of figure 3.9 a target image pixel $p_{i,j}$ is bounded by the related pixels. In this sub figure the rows are mentioned as i and columns as j .

Likewise, for our model to forecast demand in a particular hour $D_{t,j}$, which is surrounded by some historical hour values as shown in the combined hours field of Table 4.1. So, the demand $D_{t,j}$ is then relies on the combined related feature value W_t (at time t) and f_j is factor j . Proposed CNN model is then applied to this time series feature data by reflecting the relations with the image sample and thus dealing

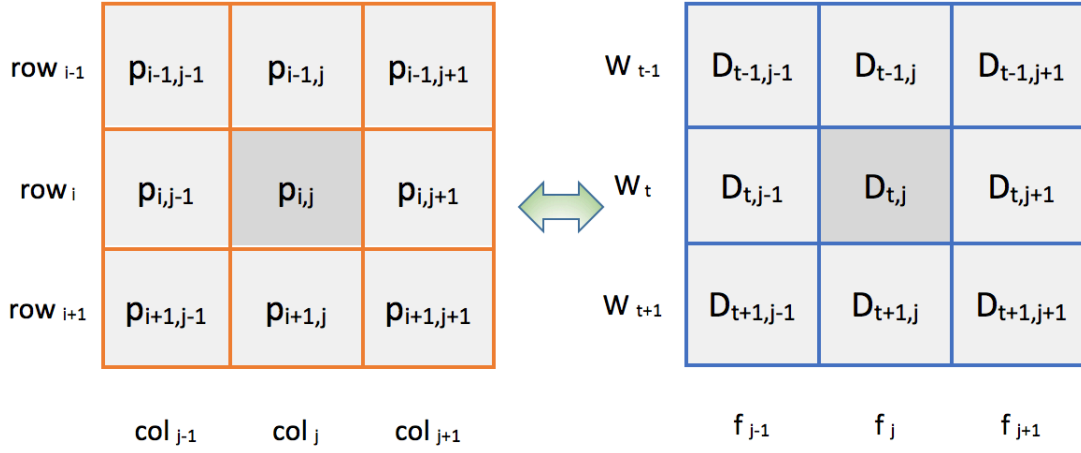


Figure 3.9. Comparing Image pixel with Bike rental demand sample

this bike rental demand forecasting as an image processing problem.

3.4.3 Model Structure

To build the forecasting model to estimate hourly rental bike demand a Convolutional Neural Network model is applied. The structure of a CNN based regression model is shown in figure 3.10.

Like any other machine learning method there are two different fragment of it, one is training and the other is testing. For the training part, the datasets gets isolated by keeping the related fields which affects the bike rental demands as features through the feature extraction methods. Afterwards, multiple structures of CNN is analyzed in order to find the best structure of the model. Three different structures of CNN is applied with different convolutional layer count shown in figure 3.11, figure 3.12 and figure 3.13. In the next section, this model is implemented on the testing set to forecast the hourly rental demand and then evaluated with four different evaluation matrices to calculate their effectiveness.

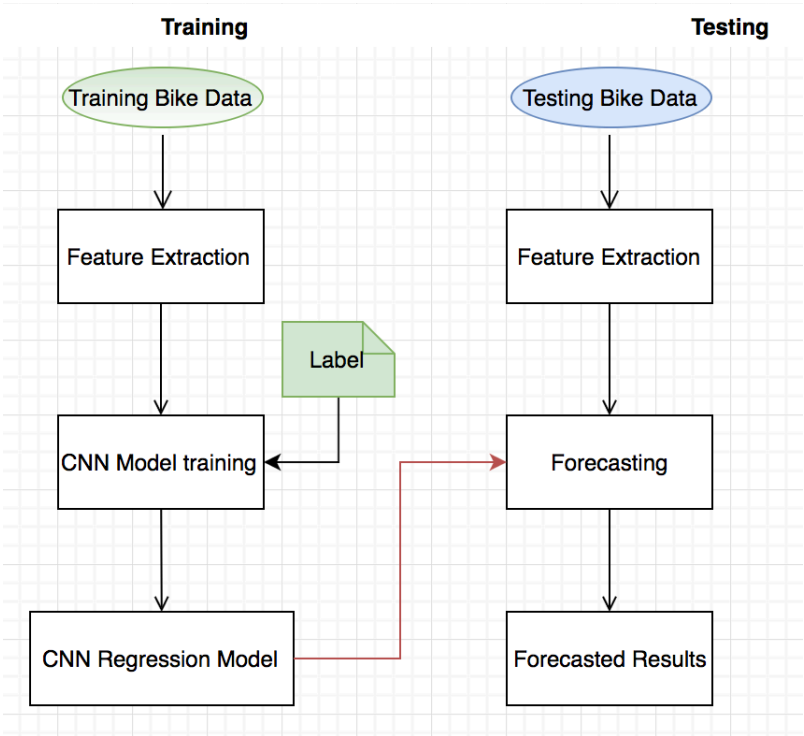


Figure 3.10. Structure of a CNN based regression model

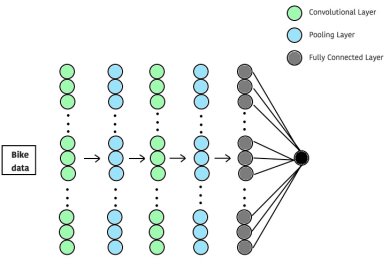


Figure 3.11. 5 Layer CNN

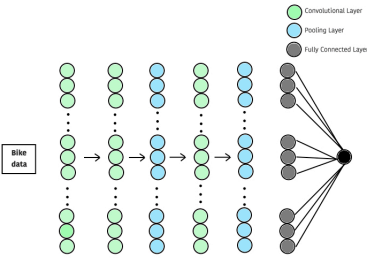


Figure 3.12. 6 Layer CNN

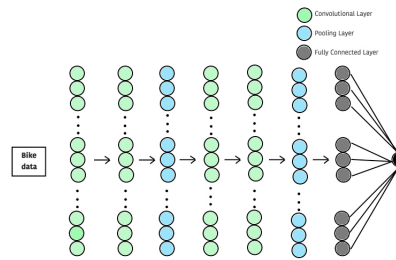


Figure 3.13. 7 Layer CNN

In figure 3.11 there are 5 layers in that CNN structure where, one convolutional layer is connected to one pooling layer which is then connected to another convolutional layer and then again to another pooling layer and finally to a fully connected layer. In figure 3.12 there is one extra convolutional layer and in figure 3.13 two extra convolutional layer from the 5 layer CNN and thus constructs 6 Layer CNN and 7 Layer CNN respectively.

CHAPTER 4

EXPERIMENT & RESULTS

4.1 Dataset Combination & Feature Selection

In order to apply machine learning algorithm to forecast bike rental demand effectively, important features from the datasets (City bike trip data, Weather data, Holiday data) needs to be collected. From the target dataset mentioned earlier in this thesis, the common fields like ‘year’, ‘month’, ‘day’, ‘hour’ counted once. Also those bike data contains more than 3million rental data, male data is about 2,342,359 and female data is about 651,398 and rest is unknown gender data. These bike rental data is then combined with the weather and holiday data as shown in Table 4.1. ‘Feature no.’ field shows the number of features related to the second column which is the name of the feature. Then Table 4.1 also shows the type of the features (numeric or binary), source of the features (1= city bike trip data, 2= weather data, 3= holiday data) and the combined hours means the hour numbers combined together to build the dataset. For example, 1-8 feature is about the hourly rental count, which is a numeric data, for combined hours field hour -1 is the previous hour data, -24 is the data from 24 hours earlier and so on.

4.2 Baseline Algorithms

In order to compare the results of proposed method some baseline models are chosen. These are- Linear regression, Neural network, Support vector regression (SVR) which are very popular and very effective machine learning algorithms in recent time. A brief description about this algorithms are given below in order to understand the methods better and how it compares to our model.

Table 4.1. Feature selection

Feature no.	Name	Type	Combined hours	Source
1-8	Hourly rental count	numeric	-1,-24,-48,-72,-96,-120,-144,-168	1
9-16	Age	numeric	-1,-24,-48,-72,-96,-120,-144,-168	1
17-24	Trip duration	numeric	-1,-24,-48,-72,-96,-120,-144,-168	1
25-32	Bike count	numeric	-1,-24,-48,-72,-96,-120,-144,-168	1
33-40	Month	numeric	-1,-24,-48,-72,-96,-120,-144,-168	1
41-48	Day	numeric	-1,-24,-48,-72,-96,-120,-144,-168	1
49-56	Snow depth	numeric	-1,-24,-48,-72,-96,-120,-144,-168	2
57-64	Snow fall	numeric	-1,-24,-48,-72,-96,-120,-144,-168	2
65-72	Wind speed	numeric	-1,-24,-48,-72,-96,-120,-144,-168	2
73-80	Temperature	numeric	-1,-24,-48,-72,-96,-120,-144,-168	2
81-88	Precipitation	numeric	-1,-24,-48,-72,-96,-120,-144,-168	2
89-96	Holiday	binary	-1,-24,-48,-72,-96,-120,-144,-168	3
97-104	Weekday	binary	-1,-24,-48,-72,-96,-120,-144,-168	3
105-112	Weekend (not holiday)	binary	-1,-24,-48,-72,-96,-120,-144,-168	3

4.2.1 Linear Regression

Regression is one of the most widely used models in recent years according to [56]. This algorithm assumes that the target variable y on features $x_1, x_2, x_3 \dots x_n$ is linear, thus it tries to find the best suitable model to fit the target. This model is then used to do prediction [57].

From the figure 4.1 we can see a simple example of linear regression. For two variable x & y and the relationship between them is $y = ax + b$. From this example at any given value of x , dependent variable y can be predicted. But many lines can fit the model perfectly, for which residuals should be taken into account. Suppose if, $A(x_1, y_1)$ using $y = ax + b$ can predict $A'(x_1, ax_1 + b)$ then the disparity between A and A' is called residual [58]. The residual sum of squares, $RSS = (y_1 - ax_1 - b)^2 + (y_2 - ax_2 - b)^2 + \dots + (y_n - ax_n - b)^2$ [figure 4.2].

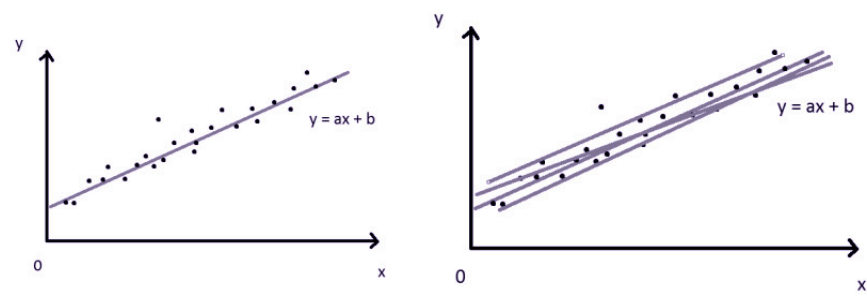


Figure 4.1. Simple Linear Regression [3]

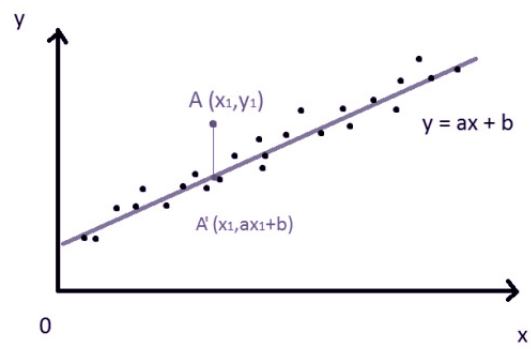


Figure 4.2. Residual example [3]

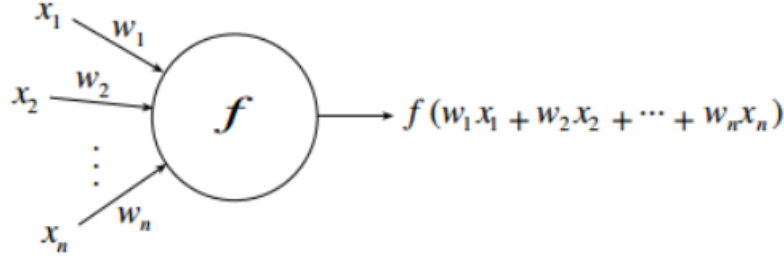


Figure 4.3. Abstract neuron [4]

4.2.2 Neural Network

Neural network is currently one of the most discussed and most commonly used machine learning algorithm. It is getting more and more attention from the researchers all over the world [4]. Its methodology is based on animal neurons, with a joined assembly of simple elements. The internal connection strengths ensures the processing ability of the network or weights which is obtained through learning [59].

According to [4] an abstract neuron [figure 4.3] can have n inputs. Each input frequency t can convey x_t information, weight w_t and f being the primitive function. The transmitted information is integrated at the neuron and the primitive function is then evaluated [4].

Neural network has the ability to combine n basic functions. In figure 4.4 four functions f_1, f_2, f_3, f_4 , network function Φ is estimated at the point x, y, z . Weights $\alpha_1, \alpha_2, \alpha_5$ can also create different network functions.

4.2.3 Support Vector Regression

Support vector machine (SVM) is a very powerful and widespread supervised machine learning model [60] which are used to solve regression or classification prob-

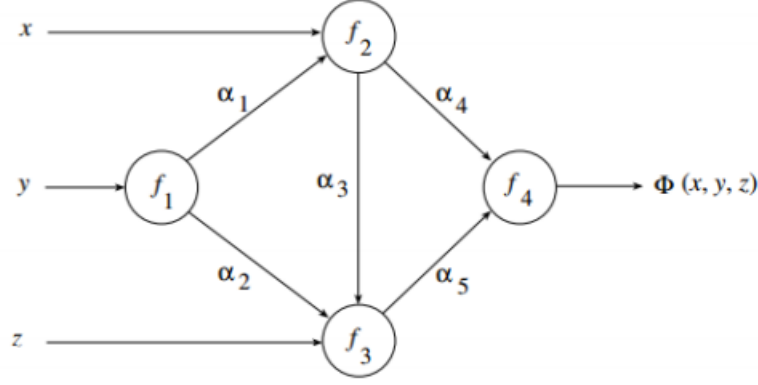


Figure 4.4. Functional model of a Neural Network [4]

lems. The version of it that deals with regression is known as Support Vector Regression (SVR). SVR is commonly known for its high level of generalization, therefore, this model can be used in new, previously unobserved data with high accuracy. In SVR figure 4.4, support vectors fabricated on the ϵ tube bounding decision shell, called training samples. Residuals which are fewer than ϵ has no influence on predictions.

Suppose, for a given dataset $(X_i, Y_i)_{i=1}^{i=N}$, where input variables X and output Y , SVR estimates the relationship as :

$Y = W \cdot \Phi(X) + b$, here $\Phi(X)$ is a non-linear kernel function which non-linearly maps from the input space X to the feature space. By minimizing the following function coefficients W and b are determined :

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{N} \sum_{i=1}^N \xi_i + \xi_i^*$$

with constraints

$$Y_i - W \cdot \Phi(X_i) - b \leq \epsilon + \xi_i$$

$$W \cdot \Phi(X_i) + b - Y_i \leq \epsilon + \xi_i^*$$

$$\xi, \xi_i^* \leq 0$$

Here, to attain acceptable results W the weight should be as flat as achievable.

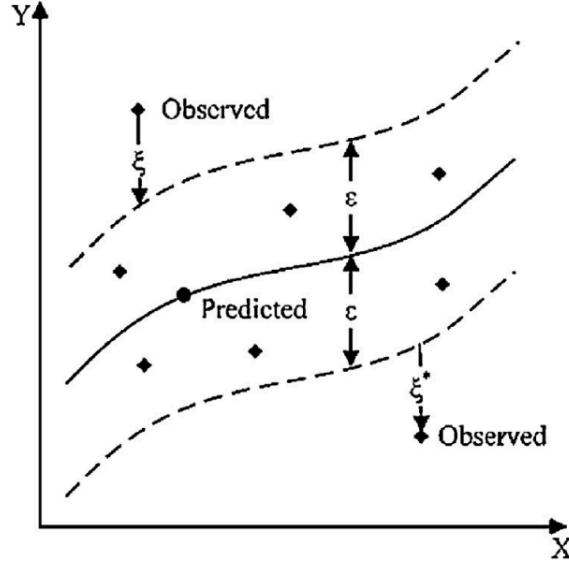


Figure 4.5. Non-linear SVR [5]

ξ_i ξ_i^* are the residuals beyond ϵ and the regularization parameter is the cost C .

4.3 Evaluation Metrics

In order to assess the model accuracy, this work uses four evaluation metrics. Mean absolute percentage of error (MAPE), Root-mean-square error (RMSE), normalized Mean absolute error (NMAE), and Normalized root-mean-square error (NRMSE).

MAPE metric has been used in a amount of prediction studies [61] [62]. It describes the average absolute error as percentage by the following equation-

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \right)$$

Here, y_i is the actual value, \hat{y}_i is the predicted value and N being the number of observations. The value of MAPE relies on the score and the lower the score is higher the performance is. MAPE value for lower forecast can not exceed 100%, but there is no limit for higher forecasts. MAPE can also not be used if there are zero

values as it will be divided by zero.

RMSE depends on the measure of the variables and should be used to compare forecasting for the similar series across the models as relative measure. Similar to MAPE, the smaller the error better the forecasting.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

One issue with evaluating with RMSE is the forecast error varies across time. Non-linearity in the model or deviations in exogenous variables can cause this issue. According to [63] no severe statistical analysis can be put on the RMSE because they are not approximations of any factor in the model.

NMAE measure normalizes the mean absolute error by the range of possible data values. NMAE is defined as follows-

$$NMAE = \frac{\sum_{i=1}^N (|\hat{y}_i - y_i|)}{\sum_{i=1}^N (|y_i|)}$$

NRMSE is the normalized root mean square error which compares between models with different scales. Similar to RMSE the value is expressed as percentage and lower value means less residual differences.

$$NRMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i)^2}}$$

4.4 Results

This section focuses on how the proposed CNN model is actually performing against the three baseline models (Linear regression, Neural Network, SVR). As mentioned earlier Table 4.1 we selected the features from those three datasets and created the time series featureset. The featureset contains 9×14 , total 126 features which has current hour data as well as 8 previous hours data based on the Table 4.1. Two different training and testing dataset were created based on monthly bike rental demand records. These dataset counts are also shown in Table 4.2. Proposed model

Table 4.2. Dataset counts

Training set		Testing set	
Male	Female	Male	Female
3612	3529	720	714

CNN is used from Tensorflow with the parameter estimation of AdamOptimizer, filter 3×3 , max pooling size 2×2 .

4.5 Discussion

From the results in Table 4.3 and Table 4.4 we can see the comparison between methods for forecasting bike rental demand for both female and male. Comparing CNN with other methods for both genders shows that CNN outperforms other baseline models with significant advantages constantly in all four different metrics. It also shows that results gained from the male datasets are far superior than the result gained from female datasets. MAPE results gained for female varies from 8-12 % for baseline models where CNN shows 7 %. It also shows better percentage level for male dataset where CNN is only 3-4 %. The reason for male data to show better forecasting results is because there are a lot of data available for male users as shown in the figure 3.5. As the results from those two table showing CNN is performing better than the baseline models it is understandable that the goal of the thesis and the methodology used to achieve this, is proving to be effective.

Even though CNN results are better than baseline models it is still confusing to decide which structure of CNN should be used. For female dataset 5 layer CNN is working better than others where in male dataset 7 layer CNN is working better. In overall, treating this bike rental data as an image processing sample and doing demand forecasting of hourly rental bike demand using CNN is a successful demand

Table 4.3. Comparison of results between proposed and baseline models (Female dataset)

Model Name	Dataset Used - Female			
	MAPE	NMAE	RMSE	NRMSE
Linear Regression	8.987	0.062	0.509	0.084
Neural Network	11.263	0.084	0.677	0.112
SVR	8.869	0.084	0.052	0.085
5 Layer CNN	7.373	0.051	0.327	0.045
6 Layer CNN	7.654	0.050	0.043	0.073
7 Layer CNN	7.489	0.050	0.425	0.070

Table 4.4. Comparison of results between proposed and baseline models (Male dataset)

Model Name	Dataset Used - Male			
	MAPE	NMAE	RMSE	NRMSE
Linear Regression	4.330	0.038	0.378	0.052
Neural Network	6.068	0.053	0.548	0.075
SVR	4.529	0.039	0.408	0.056
5 Layer CNN	4.059	0.036	0.361	0.049
6 Layer CNN	4.031	0.036	0.347	0.048
7 Layer CNN	3.531	0.030	0.327	0.045

forecasting technique.

CHAPTER 5

CONCLUSION & FUTURE WORK

This chapter will review the whole thesis and summarize the current need of demand forecasting and the approach of bike sharing service demand forecasting. Few future work and research will also be mentioned.

5.1 Conclusion

In the years to come “Big Data” term will continue to grow bigger and bigger, which makes it essential for companies to embrace this challenge with open arms and learning, implementing the required skills to tackle this. This thesis, which is mainly focuses on the demand forecasting in different fields, first recognized few challenges it poses and showing potential of using it for the benefit of organizations. It also a must to dedicate enough time and resources to be able to overcome the issues.

This thesis is specially looking at a particular transportation demand forecasting by doing hourly bike rental demand forecasting where the idea of the topic came from winning the Big data competition where the main goal was to improve the bike share program (as mentioned earlier). In order to do the bike rental demand forecasting first we did some statistical analysis on the bike rental data. From those analyses it is found that there is a big prospective of trying to attract more female users in to this bike share program. Also the visualization of the different factors of the data also showed customers mainly uses bike share program for doing short trips which makes it very important to have rentable bikes available or free parking spaces in the stations whenever customer wants. For the first problem, different types of bikes according to gender preference and forecasting hourly rental bike demand using CNN

is for the latter problem is proposed. The main contribution of this thesis is from the results gained from the experiment, which proves that doing forecasting using CNN and treating bike samples as image samples to solve regression problem is a very reasonable and achievable way to deal with the problems stated above. Nevertheless, CNN performs better than the baseline models but it is not significant. Reason being the lack of datasets may reduce the performance of the overall model.

5.2 Future Work

There are many additional work can be done in this topic as demand forecasting is proving to be one of the most important area for many organizations.

In future a demand forecasting experiment on the health care program can be carried out as it is also a very important section of demand forecasting.

For the transportation demand forecasting in bike share program, there are too many factors which may cause variance in user's behavior such as - biking events or marathon, extremely snowy day or tornado, government restriction regarding traffic. Therefore, it is almost impossible to consider all the factors in one research, thus, this thesis focuses on building a model to forecast the hourly rental bike demand to predict the number of rental demand in future based on the historical data of bike rentals, weather and holiday data. So separating every docking stations and treating them individually and forecasting their rental demands can be another very useful way to solve this problem.

The results also showed inconsistency between the different layers of CNN so finding the best structure needed for the different datasets is also a very important task to work on. It is also possible to implement transfer learning on the bike rental data forecasting model from another transportation forecasting model such as taxi or uber data. Maybe doing this we can get more insights on the users behavior which

will make our model more useful.

Finally, “Big Data”, as we all know the bigger the dataset is better the performance will be. So collecting more data can only improve the forecasting results and predicting accuracy. But it is also important to clear out any noises it may have in a very tactful way to get the best possible results of demand forecasting.

REFERENCES

- [1] “Convolutional Neural Networks.” <http://deeplearning.net/tutorial/lenet.html>. [Retrieved: January, 2017].
- [2] “Holidays in New York in 2015.” <http://www.officeholidays.com/countries/usa/regional>. [Retrieved: January, 2015].
- [3] W. Wang, “Forecasting bike rental demand using new york citi bike data,” 2016.
- [4] R. Rojas, *Neural networks: a systematic introduction*. Springer Science & Business Media, 2013.
- [5] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, “Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy,” *Applied Energy*, vol. 123, pp. 168–178, 2014.
- [6] M. Hart, J. Rašner, and X. Lukoszová, “Demand forecasting significance for contemporary process management of logistics systems,” in *Congress Proceedings-CLC 2012: Carpathian Logistics Congress*, TANGER Ltd., 2012.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [8] N. M. Richards and J. H. King, “Three paradoxes of big data,” *Stan. L. Rev. Online*, vol. 66, p. 41, 2013.
- [9] P. Tucker, “My predictions for 2014 and beyond?and why they might not come true,” Oct 2013.

- [10] S. Madden, “From databases to big data,” *IEEE Internet Computing*, vol. 16, no. 3, pp. 4–6, 2012.
- [11] D. Arribas-Bel, “Accidental, open and everywhere: Emerging data sources for the understanding of cities,” *Applied Geography*, vol. 49, pp. 45–53, 2014.
- [12] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” *arXiv preprint arXiv:1511.08458*, 2015.
- [13] H. R. Varian, “Big data: New tricks for econometrics,” *Journal of Economic Perspectives*, vol. 28, pp. 3–28, May 2014.
- [14] E. S. Silva, Y. Wu, and U. Ojiako, “Developing risk management as a competitive capability,” *Strategic Change*, vol. 22, no. 5-6, pp. 281–294, 2013.
- [15] B. Brown, M. Chui, and J. Manyika, “Are you ready for the era of ?big data?,” *McKinsey Quarterly*, vol. 4, no. 1, pp. 24–35, 2011.
- [16] T. J. Kaufman, P. A. Weaver, and J. Poynter, “Success attributes of b&b operators,” *The Cornell Hotel and Restaurant Administration Quarterly*, vol. 37, no. 4, pp. 29 – 33, 1996.
- [17] A. Walker, “Trends in big data: A forecast for 2014.”
- [18] C. Duhigg, “How companies learn your secrets,” *The New York Times*, vol. 16, p. 2012, 2012.
- [19] T. Bacon, “Big bang? when ’big data’ gets too big — travel industry news & conferences - eyefortravel.”
- [20] N. Silver, *The signal and the noise: the art and science of prediction*. Penguin UK, 2012.
- [21] G. West, “Big data needs a big theory to go with it,” Apr 2013.

- [22] C. Wesson and J. Cottier, “Big sites, big questions, big data, big problems: Scales of investigation and changing perceptions of archaeological practice in the southeastern united states,” *Bulletin of the History of Archaeology*, vol. 24, 2014.
- [23] A. Skupin and P. Agarwal, “1 introduction: What is a self-organizing map?,” 2008.
- [24] L. Einav and J. Levin, “The data revolution and economic analysis,” *Innovation Policy and the Economy*, vol. 14, no. 1, pp. 1–24, 2014.
- [25] H. Hassani, S. Heravi, and A. Zhigljavsky, “Forecasting european industrial production with singular spectrum analysis,” *International journal of forecasting*, vol. 25, no. 1, pp. 103–118, 2009.
- [26] H. Hassani, A. Webster, E. S. Silva, and S. Heravi, “Forecasting us tourist arrivals using optimal singular spectrum analysis,” *Tourism Management*, vol. 46, pp. 322–335, 2015.
- [27] J. Needham, *Disruptive possibilities: how big data changes everything.* ” O’Reilly Media, Inc.”, 2013.
- [28] R. Hyndman and G. Athanasopoulos, “Forecasting: Principles and practice, section 7/7,” *URL: <http://otexts.com/fpp/7/7>. Accessed on*, vol. 10, no. 13, p. 2015, 2013.
- [29] D. Jadhav, “Big data: the new challenges in data mining,” *Int J Innov Res ComputSci & Technol*, vol. 1, no. 2, pp. 39–42, 2013.
- [30] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable realtime data systems.* Manning Publications Co., 2015.
- [31] S. Lohr, “The age of big data,” *New York Times*, vol. 11, no. 2012, 2012.

- [32] H. T. Nguyen and I. T. Nabney, “Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models,” *Energy*, vol. 35, no. 9, pp. 3674–3685, 2010.
- [33] T. Callahan, D. Guzman, and M. Sumeren, “Effective demand forecasting in the health care supply chain,” 2004.
- [34] P. Jensen, J.-B. Rouquier, N. Ovtracht, and C. Robardet, “Characterizing the speed and paths of shared bicycle use in lyon,” *Transportation research part D: transport and environment*, vol. 15, no. 8, pp. 522–524, 2010.
- [35] P. DeMaio, “Bike-sharing: History, impacts, models of provision, and future,” *Journal of Public Transportation*, vol. 12, no. 4, p. 3, 2009.
- [36] S. Shaheen, S. Guzman, and H. Zhang, “Bikesharing in europe, the americas, and asia: past, present, and future,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2143, pp. 159–167, 2010.
- [37] J. Moura, A. Ibeas, *et al.*, “Implementing bike-sharing systems,” in *Proceedings of the Institution of Civil Engineers-Municipal Engineer*, vol. 164, 2011.
- [38] L. M. Martinez, L. Caetano, T. Eiró, and F. Cruz, “An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of lisbon,” *Procedia-Social and Behavioral Sciences*, vol. 54, pp. 513–524, 2012.
- [39] V. P. Kumar and M. Bierlaire, “Optimizing locations for a vehicle sharing system,” in *Swiss Transport Research Conference.(Cited on pages 2, 8, and 24.)*, 2012.
- [40] J.-R. Lin and T.-H. Yang, “Strategic design of public bicycle sharing systems with service level constraints,” *Transportation research part E: logistics and*

- transportation review*, vol. 47, no. 2, pp. 284–294, 2011.
- [41] T. Raviv, M. Tzur, and I. A. Forma, “Static repositioning in a bike-sharing system: models and solution approaches,” *EURO Journal on Transportation and Logistics*, vol. 2, no. 3, pp. 187–229, 2013.
 - [42] R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Bušić, “Large-scale vehicle sharing systems: analysis of vélib’,” *International Journal of Sustainable Transportation*, vol. 7, no. 1, pp. 85–106, 2013.
 - [43] E. Miller-Hooks and R. Nair, “Fleet management for vehicle sharing operations,” tech. rep., 2010.
 - [44] F. Leurent, “Modelling a vehicle-sharing station as a dual waiting system: stochastic framework and stationary analysis,” 2012.
 - [45] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, “Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system,” *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 455–466, 2010.
 - [46] R. Rixey, “Station-level forecasting of bikesharing ridership: Station network effects in three us systems,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2387, pp. 46–55, 2013.
 - [47] P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury, “Shared bicycles in a city: A signal processing and data analysis perspective,” *Advances in Complex Systems*, vol. 14, no. 03, pp. 415–438, 2011.
 - [48] J. Froehlich, J. Neumann, and N. Oliver, “Measuring the pulse of the city through shared bicycle programs,” *Proc. of UrbanSense08*, pp. 16–20, 2008.

- [49] N. Lathia, S. Ahmed, and L. Capra, “Measuring the impact of opening the london shared bicycle scheme to casual users,” *Transportation research part C: emerging technologies*, vol. 22, pp. 88–102, 2012.
- [50] P. Borgnat, P. Abry, P. Flandrin, and J.-B. Rouquier, “Studying lyon’s vélo’v: a statistical cyclic model,” in *ECCS’09*, Complex System Society, 2009.
- [51] R. Giot and R. Cherrier, “Predicting bikeshare system usage up to one day ahead,” in *Computational intelligence in vehicles and transportation systems (CIVTS), 2014 IEEE symposium on*, pp. 22–29, IEEE, 2014.
- [52] J. Shu, M. Chou, Q. Liu, C.-P. Teo, and I.-L. Wang, “Bicycle-sharing system: deployment, utilization and the value of re-distribution,” *National University of Singapore-NUS Business School, Singapore*, 2010.
- [53] D. Chemla, F. Meunier, and R. W. Calvo, “Bike sharing systems: Solving the static rebalancing problem,” *Discrete Optimization*, vol. 10, no. 2, pp. 120–146, 2013.
- [54] J.-H. Lin and T.-C. Chou, “A geo-aware and vrp-based public bicycle redistribution system,” *International Journal of Vehicular Technology*, vol. 2012, 2012.
- [55] J. W. Yoon, F. Pinelli, and F. Calabrese, “Cityride: a predictive bike sharing journey advisor,” in *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on*, pp. 306–311, IEEE, 2012.
- [56] G. A. Seber and A. J. Lee, *Linear regression analysis*, vol. 936. John Wiley & Sons, 2012.
- [57] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

- [58] D. Montgomery, E. Peck, and G. Vining, “Wiley series in probability and mathematical statistics,” *Introduction to linear regression analysis [J]*, 2012.
- [59] K. Gurney, *An introduction to neural networks*. CRC press, 1997.
- [60] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics Springer, Berlin, 2001.
- [61] Z. Hu, Y. Bao, R. Chiong, and T. Xiong, “Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection,” *Energy*, vol. 84, pp. 419–431, 2015.
- [62] R. E. Edwards, J. New, and L. E. Parker, “Predicting future hourly residential electrical consumption: A machine learning case study,” *Energy and Buildings*, vol. 49, pp. 591–603, 2012.
- [63] R. C. Fair and R. J. Shiller, “Comparing information in forecasts from econometric models,” *The American Economic Review*, pp. 375–389, 1990.

CURRICULUM VITA

Safat Mahmood

Prairie View A&M University
 Department of Electrical and Com-
 puter Engineering
 P.O. Box 519, Prairie View, 77446

Phone: (713) 377-8388
 Email: safatmahmood@gmail.com

Education

M.S. Electrical and Computer Engineering, Prairie View A&M University, 2015.

B.Eng. Electrical/Electronics Engineering, Queen Mary University of London, 2012.

Employment

Prairie View A&M University, Research Assitant, 2015–2016.

Pacific Jeans Ltd., Assistant Engineer, 2013-14.

Professional Skills

Solid foundation of theoretical knowledge, with sound practical electrical engineering/ programming skills, and strong/quick learning ability.

Natural Language Processing (NLP), Information Extraction & sentiment analysis, Machine learning, Deep learning, Network and Cyber Security, Robotic Sensing, Semiconductors, Microcontrollers

Embedded system, assembly language , FPGA, C++, python, Matlab, HTML, CSS and many more.

Achievements & Social works

Won First Prize in the 2016 IEEE Big data Competition in Chengdu, China.

Became Outstanding MSEE student of the department of Electrical and Computer Engineering, Prairie View A&M University, 2016-17

Got sponsored by British Airways for my Bachelors final year project.

Helped flood victims of Bangladesh from school.

Completed my jury services for Southwark crown court, London.