IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 2, February 2019

Prediction of Bike Rentals

Dr. Shanthi Mahesh[1], Aditya Warrier[2], Deepthi P[3], Dimple K H[4]

HOD, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India[1]

Student, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India[2,3,4]

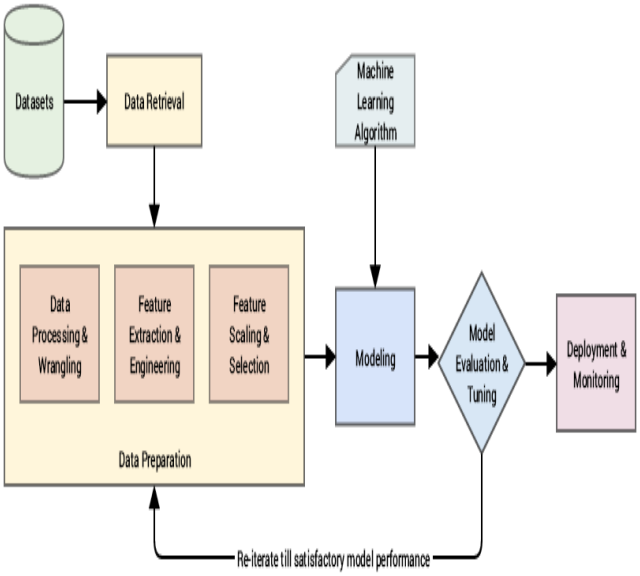
Abstract: Bike rental predictions forecasts the demand for bikes rentals in dependency of weather conditions like the temperature and calendric information e.g. holidays. To make predictions machine learning is used. The dataset contains some time features which we translate to indicator variables, as well as weather information for that day. Generally in bike rental systems it is very important that the administrators should know how many bikes are to be placed in each station, knowing this count enables them in arranging the required number of bikes at the stations and decide whether a particular station needs to have extra number of bikes or not. So in this research work we study prediction associations to enhance their administrations and items in view of clients' input.

Keywords: Random Forest, Neural Network, Prediction, Result Analysis.

I. INTRODUCTION

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns. They interpret sensory data through a kind of machine perception, labeling or clustering raw input. The patterns they recognize are numerical, contained in vectors, into which all real-world data, be it images, sound, text or time series, must be translated. Neural networks help us cluster and classify.

II. GENERAL STEPS FOLLOWED IN PREDICTION PROJECTS



The above figure shows the general steps followed in Prediction project. Where initially data is retrieved. Which is then pre-processed, scaled and selected after extraction and fed into the model to obtain the predictions. The accuracy and predictions can vary based on many factors such as pre-processing techniques, the algorithm used, data been fed into the model as well as the model training procedures.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

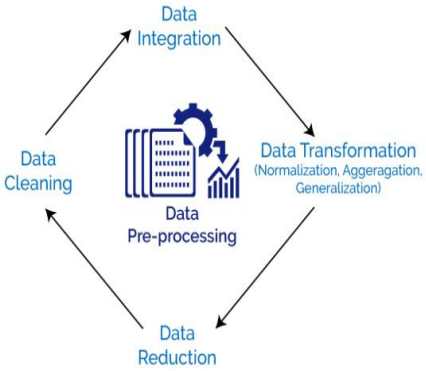
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 2, February 2019

III. RELATED SURVEY

Past research on pre-processing of data

Pre-processing involves segregating the data into different parameters such as time of day, date, holiday etc. The dataset was then subjected to a multiple regression model and parallel to Random Forrester approach. The multiple regression model was based on 7 parameter (weather and seasonal) upon which the bike rental demand depended, respectively weather, temp, attempt, humidity, wind speed, date time and the season [1].

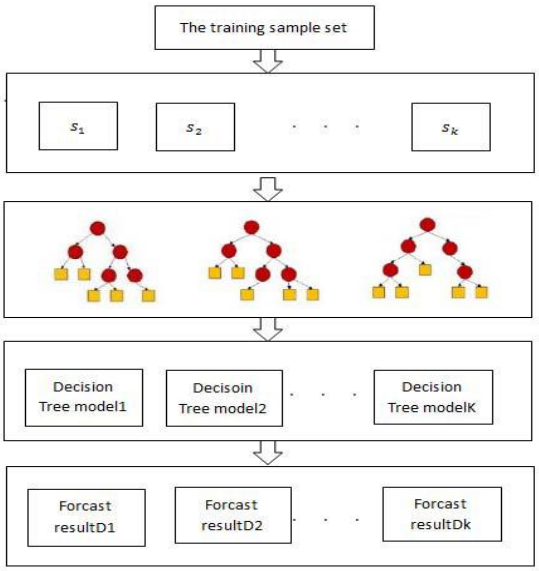


In the above figure shows the pre-processing of data having different steps as dependable factors.

Each training sample consists of date and hour, season, holiday, working-day, weather, temp, atemp (the temperature it feels like), humidity, and wind speed. Date and hour are provided as a single string, weather is a categorical variable with 4 levels representing different weather conditions, holiday is a binary indicator variable representing whether the particular day was a holiday. [3]. Some of the attributes of the dataset are date, time, season, holiday etc. The pre-processing is done to reduce the noise and any inconsistency. Feature engineering for adding some new features such as dividing temperature, humidity into different categories for a cleaner dataset. [2]

Past research on Prediction using Random Forest approach

The Random Forrest approach included the use of decision tree pruning, a kind of typical single classifier, the first step of training set is to recursive analysis, generate a shape such as inverted tree structure; The second step analysis of the tree from the root node to leaf node path, produce a series of rules. Finally, according to these rules, classification or projections for new data. Every tree in the forest depend on a random vector, vector in the forest are all independent identically distributed. [1].



Random Forest is a meta-algorithm that combines a large number of decision-tree models, each individually built on bootstrapped samples of the data. This process of sampling the data and combining the individual decision-trees is called Bagging, and is able to reduce the variance of the predictions without increasing the bias. The predictions are

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

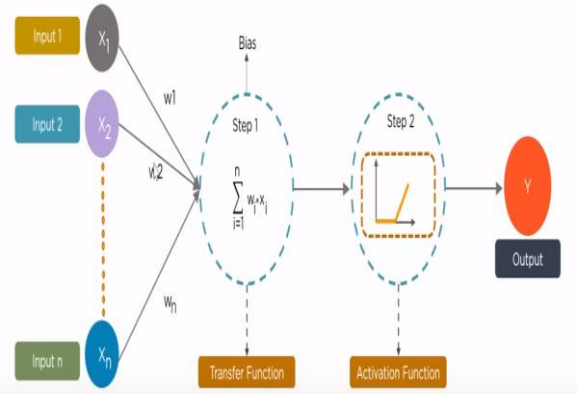
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 2, February 2019

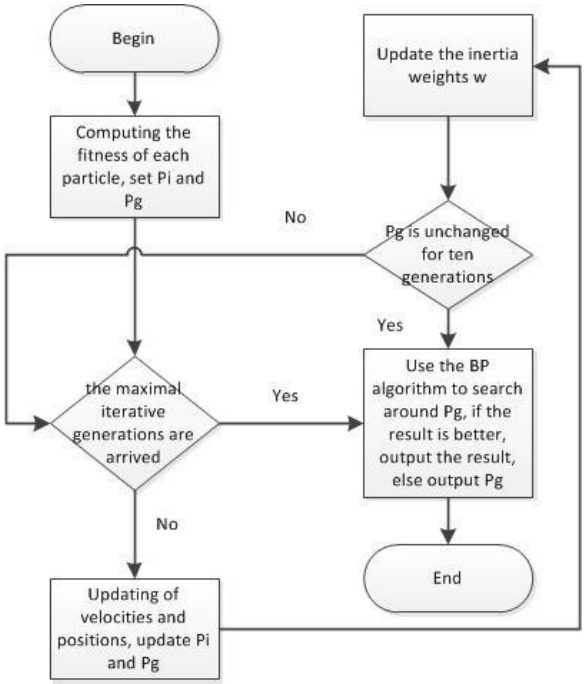
formed by taking the mean of the individual decision-tree predictions. [3] The random forest method along with enhancement with TuneRF is used [2].

Past research on Prediction using Neural Network approach.

Neural network model can realize nonlinear mapping relationship well by training a large amount of input-output dataset [4].



APSO-BP neural network model which uses APSO algorithm for global optimization has advantages of simple implementation and fast convergence speed. At the beginning of training neural network, it can find out the optimum solution very quickly. APSO-BP algorithm is better than PSO and BP at convergence speed and accuracy [4].



In APSO algorithm, inertia weight is decreased continuously in the iterative process. In the searching process of the particle swarm optimization algorithm, the search space is reduced gradually in the process of iteration. So APSO is more effective. [4] The advantage of the usage of neural networks for prediction is that they are able to learn from examples only and that after their learning is finished, they are able to catch hidden and strongly non-linear dependencies, even when there is a significant noise in the training set. The disadvantage is that NNs can learn the dependency valid in a certain period only. The error of prediction cannot be generally estimated.

Past research done on Prediction using Multiple Linear Regression

Multiple Regression is a technique where we use the dependable variables to learn a model that enables you to predict the value of the response variable, given a new record where you only know the values of the dependent variables. The

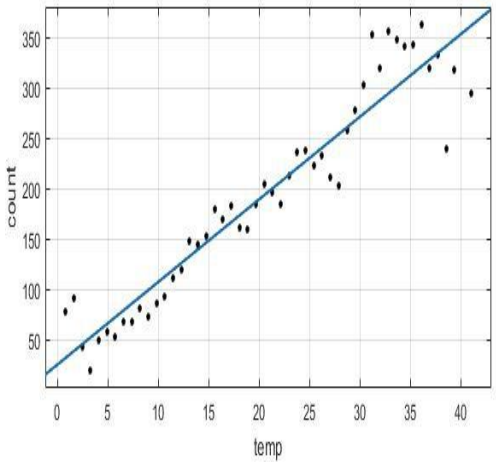
IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 2, February 2019

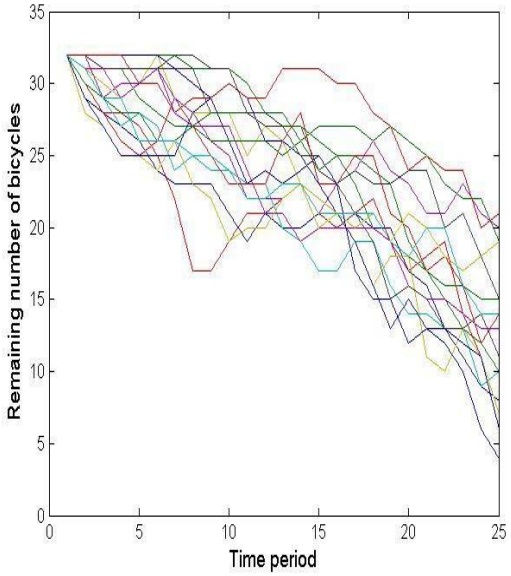
regression model established introduces the data to test required by the multiple regression analysis normality and linear relationship between the two premise condition which is satisfied. [1]



The above graph, [1] shows the relationship of the dependable variable which is the temperature mainly on the count of the number of bikes predicted.

Past research done on Prediction using K-Means Cluster Analysis

Clustering analysis can be done on the basis of features where we try to find subgroups of samples based on features or on the basis of samples where we try to find subgroups of features based on samples**.** K is an input to the algorithm for predictive analysis; it stands for the number of groupings that the algorithm must extract from a dataset, expressed algebraically as *k*. A K-means algorithm divides a given dataset into *k* clusters. The goal of K-means is to minimize the sum of the squared error over all K clusters [4]. A representative of a cluster is the mathematical mean (average) of all items that belong to the same cluster. This representative is also called a cluster centroid. K-means, which is a greedy algorithm, may converge to a local minimum. Different initialization can lead to different final clustering. One way to overcome the local minimum is to run the K-means algorithm, for a given K, with multiple different initial partitions and choose the partition with the smallest squared error[4].



The above figure denotes the results of K-means algorithm classification when taken K= 4. [4] K means algorithm is good in capturing structure of the data if clusters have a spherical-like shape. It always try to construct a nice spherical shape around the centroid. It doesn't let data points that are far-away from each other share the same cluster even though they obviously belong to the same cluster. It suffers as the geometric shapes of clusters deviates from spherical shapes. It doesn't learn the number of clusters from the data and requires it to be pre-defined.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 2, February 2019

IV. CONCLUSION

From the knowledge we gained by referring multiple project reports and survey papers belonging to the domain of Random Forest Prediction algorithm and Neural Network, we conclude that prediction algorithm approach with neural network model is best suited for training the model when there is availability of data. It also suits for models which are meant not to be domain specific.

REFERENCES

[1]. YouLi Feng , ShanShan Wang "A Forecast for Bicycle Rental Demand Based on Random Forests and Multiple Linear Regression"
[2]. Akshay Patil, Karishma Musale and BVANSS Prabhakar Rao "Bike Share Demand Prediction using Random Forests"
[3]. Jimmy Du, Rolland He, Zhivko Zhechev presented "Forecasting Bike Rental Demand"
[4]. Fei Lin, Shihua Wang, Jian Jiang, Weidi Fan, Yong Sun "Predicting Public Bicycle Rental Number using Multi-source Data"