TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

Ideal Crop Recommendation Module

Ashwini R[1], Akhila M[2], Archana R Hegde[3], Sahana J[4], A B Rajendra[5]

[1,2,3,4,5]Department of Information Science Engineering, Vidyavardhaka College of Engineering, Mysuru, India

Article Info
Volume 82
Page Number: 4600 - 4607
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 22 January 2020*

Abstract

In this era of digital revolution, Machine learning has become one of the top most emerging technologies. The machine learning algorithms are being used in various fields and applications such as image recognition, speech recognition, classification, prediction, recommendation systems, etc. Agriculture, the backbone of Indian economy, contributes to the overall economic growth of our country. To improve the current productivity status, Ideal Crop Recommendation Module uses machine learning algorithms to guide the farmers to grow the suitable crop based on soil attributes. This system can also be used to predict the yield of certain crops given the region and season.

Keywords: *Agriculture, Machine Learning, Precision Farming, KNN, Naïve Bayes, Linear Regression*

I. INTRODUCTION

India is known as an agricultural country, which contributes to its overall economy. Agriculture determines the standard of living of more than half of the country's population. The productivity of Indian agriculture is very less when compared to world standards. The agriculture sector has hardly seen any new technological developments. Due to this, most of the farmers still follow traditional methods for farming which might lead them to choose the wrong type of crop to cultivate based on their current soil condition. This limits the potential of getting a profitable yield.

A new approach known as precision farming is introduced from last 20 years in order to improve the society's agricultural needs. Precision farming is a technique dealing with farm management. It applies information technology to make sure that the soil and crops receive exactly what they require for optimum health and maximum productivity. What this definition means is precision farming is all about doing the right thing like applying correct amount of fertilizers and water, in the right place, in the right manner, at the right time to the crop. The goal is to manage crop production and increase the yield, profit, quality, reduce waste and become eco-friendly.

In India, agriculture is being practiced for ages and hence, there is a rich collection of past agricultural data. Machine learning techniques can use this data for crop recommendation. The system gives the recommendation based on the current soil N-P-K values, pH, EC (Electrical Conductivity), rainfall. The recommendation includes the suitable crops for the given soil attributes. The system also includes a yield prediction module. This module predicts the yield for the year 2018 based on factors such as season, region and the farmer. Currently this module is trained for only paddy.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

II. LITERATURE SURVEY

The recommendations for agriculture in India are done in traditional ways, which are manual based and time consuming. To improve this, precision agriculture can be practiced. The software proposed in [1] deeply analyzes the basics of precision farming and then develops a model that would support it. This paper discusses the planning and the requirements that are needed for developing a model for precision farming. The main objective of this paper was to provide direct advisory services to even the smallest farmer. The model proposed in this paper was designed for the scenario in the State of Kerala where, the average holding size of the farm was much lower than other parts of India. Hence, only minor modifications are required for this model to be deployed elsewhere in India.

In [2] States the use of agricultural data with the data mining techniques. The techniques used here are self-organizing maps and multi-dimensional scaling to reduce the data. It was concluded that when the dataset is large self-organizing maps are suitable and Sammon's mapping is suitable when the dataset is small.

In [3] Aims to solve the crucial problems of classifier selection for ensemble learning. This work has proposed a method to select the finest set of classifiers from a pool of classifiers. Finally, it was inferred that SAD (Selection by Accuracy and diversity) was better.

In [4] Proposes a system that helps the farmers in sowing the correct seed based on the requirements of soil to increase productivity and acquire profit. It also proposes a future work of improved dataset having large number of attributes. The proposed system uses an ensemble model selection technique using majority voting.

In Paper [5] explores the significance of crop selection and also the factors that lead in its determination, like government policies and production rate are discussed and depicted. To solve the crop selection problem, a method titled, Crop Selection Method (CSM) is presented. It takes into account, factors such as the soil type, crop type, water density and weather to suggest a train of crops for a farmer to select.

In this research paper [6], they have applied descriptive analytics for finding an efficient crop yield prediction for the sugarcane crop. They have created a combined dataset using the Soil dataset, Rainfall dataset and Yield dataset, to which they have applied various supervised techniques like SVM, KNN and least squared SVM.

In [7] juxtaposes the predictive accuracy of two techniques, namely ML and linear regression techniques that are here used for crop yield prediction in 10 crop datasets. Multiple Linear Regression, ANN, M5-Prime regression trees and KNN were used for this. The model was built using the real data of an irrigation zone of Mexico. It was deduced from their studies that two methods obtained the lowest errors; viz. M5-Prime and KNN. Hence they were deemed as suitable tools for massive crop yield prediction in agricultural planning.

In paper [8], they predict the yield of crops based on the existing data using Random Forest algorithm. The model was built using the real data of Tamil Nadu and the model was tested with samples. The parameters that their dataset included were – rainfall, season, temperature, crop production, perception.

III. MACRO-NUTRIENTS IN THE SOIL

The quality of the soil is determined by its pH, EC, NPK values. 'EC' is the electrical conductivity of the soil. The 'N', 'P', 'K' symbolize the nitrogen, phosphorous and potassium content of the soil.

Soil pH is the measure of acidity or alkalinity of a soil. Soil pH can affect the plant growth in several ways. The pH range of 5.5 to 7.0 is said to be the range in which the bacteria that has a hand in the change and release of nitrogen to the soil, from the

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

organic matter and fertilizers, operate. Thus making it the optimum pH range. The plant nutrients are observed to leach faster from the soil at pH values lower than 5.5. Soil pH may also affect the availability of plant nutrients.

Soil electrical conductivity (EC) is a measure of the amount of salts in soil (salinity of soil). EC plays the role as an indicator of soil health. It affects some of the major influencers of the critical soil processes. Some of these influencers are - crop yields, suitability, plant nutrient availability, and activity of soil microorganisms.

Nitrogen is very important and needed for plant growth. Nitrogen is part of the chlorophyll molecule, hence is contingently responsible for the green color of the plants. It also participates in the process of photosynthesis. If low quantity of nitrogen is found in the soil, then plants will have slight yellowish leaves and if quantity is moderate or high, it will have greener leaves.

The amount of phosphorous present in the soil is culpable for the reproductive system and the rooting system of the plant. The growth of fruits and flowers of the plants are reliant on this value. Phosphorous is required to convert other nutrients of the soil into its usable form for the plant.

The soils potassium content determines the overall growth of a plant. It predicts the strength of the plants' roots. The growth process of the plant altogether is discerned by this value. Potassium is necessary in regulating the functioning of guard cells, which helps in the exchange of water vapour, oxygen and carbon-dioxide. If K is deficient, it stunts the plants growth and results in low yield.

IV. EXISTING SYSTEM

The current system of recommendation involves direct interaction between a farmer and an agriculture expert, wherein either of the two will have to commute to meet the other in their location. These recommendations also differ from one expert to another, which confuses the farmers. The confusion which arises could result in the loss of the crop yield of a season. The present system would take days to provide the recommendation. This is hence is a time consuming process. This system tries to bridge the gap between the agriculture and technology.

V. PROPOSED SYSTEM

There are various factors that might determine the crop yield, the nutrients present in the soil being the major one.

In our system, the user can input the various attribute values of a particular soil sample. Based on these, the soil is analyzed to test the mineral levels, and other components that will favor the growth of a particular crop/plant.

Once the soil is analyzed for its EC, Nitrogen, Phosphorous, Potassium content, its pH, the system compares them with the data stored to determine the soil-crop recommendation.

The system also has a feature where the farmers can get a prediction of yield. The system is trained with the data from 2010 – 2017. The prediction results are given for the year 2018. The prediction is based on the attributes such as region, season and crop.

Crop Recommendation Module

The proposed model takes the soil attributes as its input. Using the following algorithms:

- K-Nearest Neighbor
- Gaussian Naïve Bayes

the ideal crops are recommended.

Fig 1 shows the block diagram for the crop recommendation module.

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

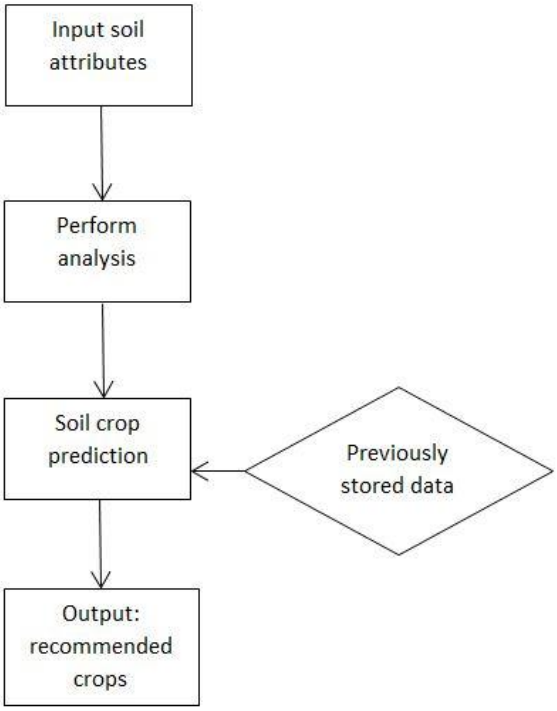

Fig 1 Block Diagram of Crop Recommendation Module

Crop Yield Prediction Module

The yield estimation for numerous crops is an important issue for agricultural planning. For achieving an effective solution to this problem Machine Learning techniques can be used. Real data of the agricultural regions of Karnataka were collected from an agriculture centre for the purpose of the same. The data we collected is only concerned with paddy crop.

Using the following Machine Learning techniques:

- Linear Regression
- K-Nearest Neighbor

the yield for the year 2018 is predicted.

Fig 2 shows the block diagram for the crop yield prediction module.

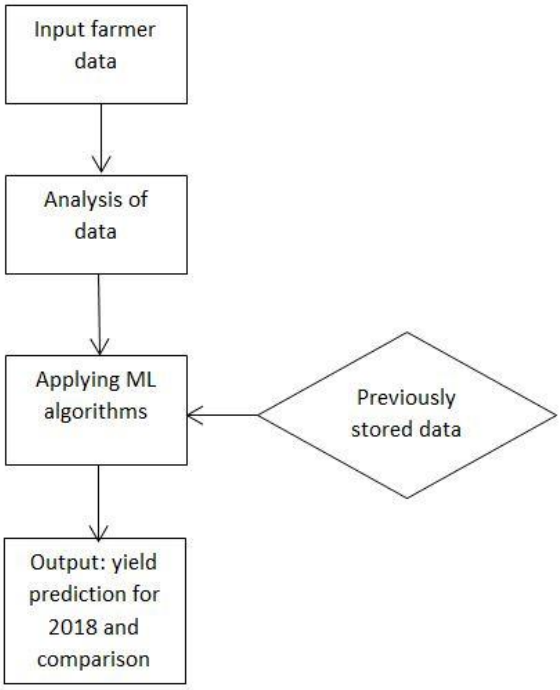

Fig 2 Block Diagram of Yield Prediction Module

VI. METHODOLOGY

Data Collection

The dataset consisting of the soil specific attributes was collected from an agricultural center in Suttur for Mysore district, Karnataka, India. Our model considers the following crops for recommendation – paddy, sugarcane, horse gram, coffee, cotton, banana, guava, mulberry, coconut. The attributes considered are – the soil pH, EC, Nitrogen, Phosphorous, Potassium, rain. The stated parameters play an important role in the crop growth.

The dataset used for yield prediction consists of the various features – the average values of the different soil attributes (pH, Rain, N, P and K) for the corresponding years from 2010 to 2017 and the yield (first harvest and second harvest) obtained, for each farmer.

Data Preprocessing

The role of data preprocessing [12] is to convert the raw data into a digestible composition. The fact is that the real world data at times is deficient - missing allocated values, missing precise attributes

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

of need, noisy-where errors exist and inconsistent-which contains dissimilarity in names.

Handling Missing Values

There might be many tuples with missing attribute values in the collected raw data. The values which are missing should be properly handled or else the inferences which are drawn will have inaccurate data results. There are 2 prominent ways of handling the missing values:

Either a row having an empty value for a distinct attribute or column should be deleted, if the missing values exceed 75%. The above method is to be used only if the dataset containing enough samples in it.

The mean, median or mode can be calculated for the feature with the replacement of missing values with them. This method is described as Leaking the data

Since, our data was collected in physical form; it was observed that many tuples had missing values when being converted to digital form. Therefore, these missing values were filled with mean of the respective attribute values.

Feature Scaling

Feature scaling [12] is usually done to limit the variable ranges so that they can be related to general framework. If there is a lot of a variation, then that feature would dominate the result. One of the most common methods is Standardization. Here, the mean is subtracted from the original value and is divided by the standard deviation for each feature.

Since we are using KNN as one of the ML techniques, the value should belong to a comparable range.

Encoding

Encoding is the transformation of categorical variables to binary or numerical counterparts. Since we are using Gaussian Naïve Bayes and KNN we aim to have all the attributes in numerical form.

Learners Used in the Model

K-Nearest Neighbors

The K-Nearest Neighbors algorithm (KNN) [9] can be used in two processes i.e., classification and regression. In both cases, training is done using k closest training examples as inputs in the feature space. The output depends on the process. In KNN regression, a discrete value for the object will be returned. This value is the mean of the values of the k nearest neighbours.

The k closest neighbours are found using the Euclidean distance. The formula for Euclidean distance between two examples x and y is given below, where n represents the number of features or attributes:

$$dist(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Gaussian Naïve Bayes

In Machine Learning, [10] Naïve Bayes learner is a simple probability model based on Bayes' Theorem. Naïve Bayes makes assumptions as to their being strong independence between features. When dealing with continuous data, we go for Gaussian Naïve Bayes.

The probability distribution of a continuous attribute v, given a class C is given by the following equation:

$$P(v|C) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(v-\mu)^2}{2\sigma^2}}$$

Where, μ is the mean of the values in v and σ is the variance of the values v associated with class C.

Linear Regression

[11] In consideration of statistics, we describe linear regression as the relationship modeling with a linear curve between the scalar response and one or more explanatory variables.The predictor functions which are linear go through modeling of relationships

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

which describe the unknown parametric model from the data. Linear regression mainly targets the conditional probability distribution of the response for the predictors' values.

VII. RESULTS

The following images show the output of our model for crop recommendation,

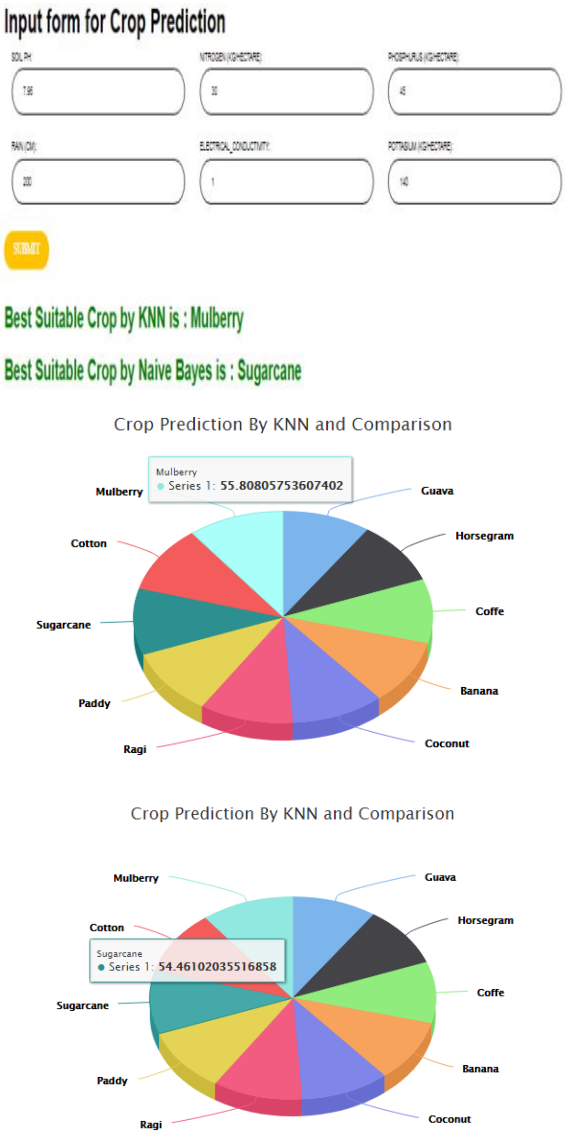

Fig 3 Crop recommendation output – 1

As shown in the above figure for the given attributes values the model recommended Mulberry and Sugarcane.

Fig 4 shows the output of the model where both the algorithms recommend the same crop for the given soil attribute values.

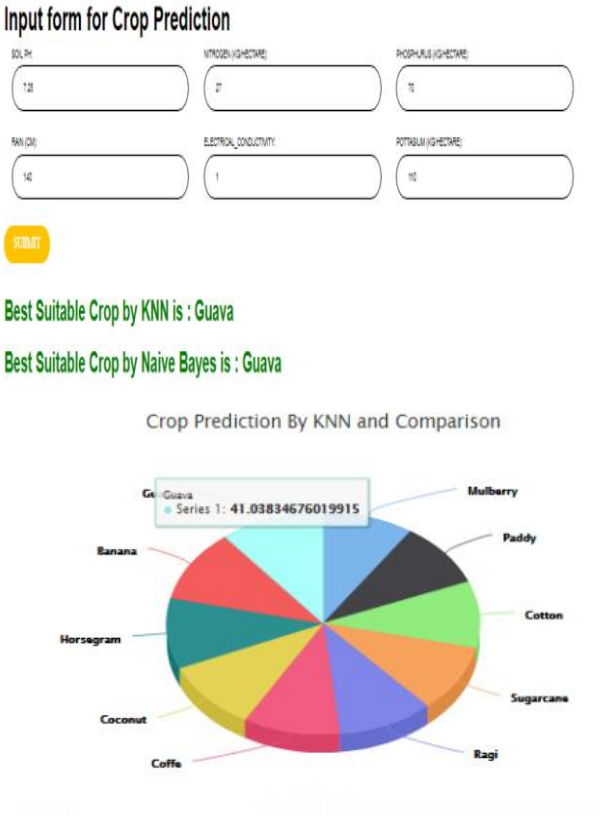

Fig 4 Crop Recommendation output – 2

The following figures show the output of our model for yield prediction,

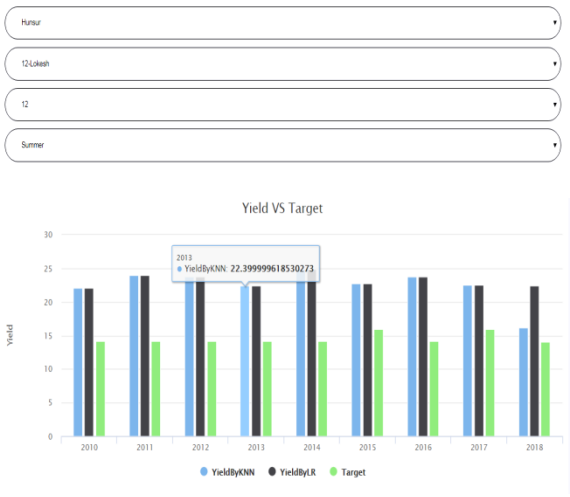

Fig 5 Yield Prediction output

The above bar graph compares the prediction of the two algorithms with the target yield.

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

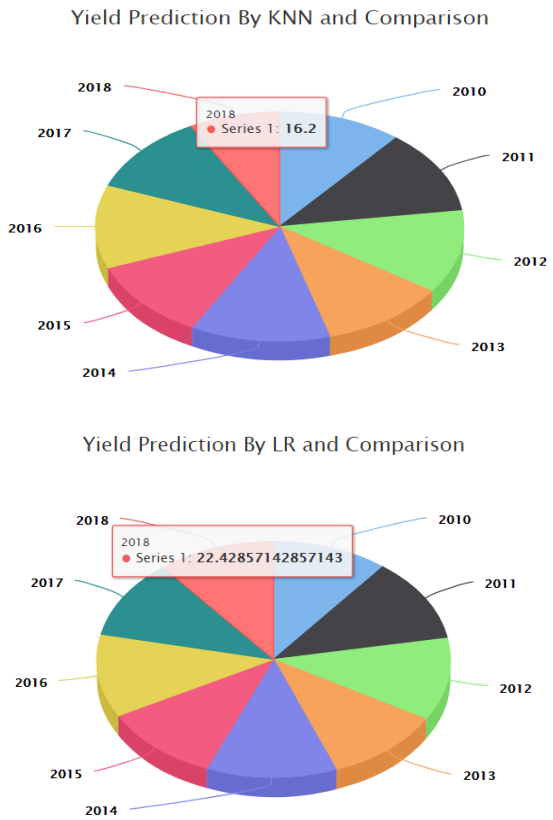

Fig 6 Pie charts for yield prediction

The above pie charts show the comparison of the two algorithms.

VIII. CONCLUSION

Agriculture plays a key role in defining the economy of a nation like India. In the prosperity of the farmers, lies the prosperity of the nation. Thus, our work would help recommending the ideal crop required for a fruitful yield. In this paper, we have summarized our study on the application of precision agriculture to improve the current status of productivity of Indian agriculture. The traditional method of crop recommendation which is currently in practice can be improved using the proposed model. Based on the literature survey conducted, the above mentioned algorithms were found to be more efficient. Hence, we have followed the same.

Our future work includes expanding the dataset to include various other features that play a role in affecting the yield of a crop. Also the model can be expanded to provide recommendation to other geographical areas as our model is bound to Mysore district. The proposed model has been trained to predict the yield for paddy crop; it can further be improved to provide the prediction to various other crops.

ACKNOWLEDGEMENTS

The manuscript is prepared by taking assistance from Accendere Knowledge Management Services Pvt. Ltd, we are thankful to them. We also express our gratitude to our teachers and mentors for guiding us throughout the work.

REFERENCES

[1] Babu, S. (2013). A software model for precision agriculture for small and marginal farmers. In 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS,) pp. 352–355.

[2] Sanghvi, Y., Gupta, H., Doshi, H., Koli, D., Ansh, A., & Gupta, U. (2015). Comparison of Self organizing maps and Sammon's mapping on agricultural datasets for precision agriculture. In 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1–5.

[3] Yang, L. (2011).Classifiers selection for ensemble learning based on accuracy and diversity.Procedia Engineering, 15, pp. 4266–4270.

[4] Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., &Nisha, J. (2017).Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32–36.

[5] Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015).Crop Selection Method to maximize crop yield rate using machine learning technique.In 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 138–145.

[6] Kumar, A., Kumar, N., & Vats, V. (2008).Efficient crop yield prediction using machine learning algorithms. International Research Journal of Engineering and Technology, 5(6), pp. 3151–3159.

January - February 2020
ISSN: 0193 - 4120 Page No. 4600 - 4607

[7] Alberto, G.-S., Frausto-Solis, J., & Ojeda-Bustamante, W. (2003).Predictive ability of machine learning methods for massive crop yield prediction. Spanish Journal of Agricultural Research

[8] Priya, P., Muthaiah, U., &Balamurugan, & M. (2018).Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Sciences & Research Technology, 7(4), pp. 1–7.

[9] k-nearest neighbors algorithm. (2019). Retrieved May 3, 2019, from https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[10] Naive Bayes classifier. (2019). Retrieved May 3, 2019, from https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[11] Linear regression. (2019). Retrieved May 3, 2019, from https://en.wikipedia.org/wiki/Linear_regression

[12] Mohit Sharma. (2018). What Steps should one take while doing Data Preprocessing? | Bitcoin Insider. Retrieved May 3, 2019, from https://www.bitcoininsider.org/article/33720/what-steps-should-one-take-while-doing-data-preprocessing