Climate Change Analysis Using Machine Learning

Himanshu Vishwakarma

Student, Department Of Computer Science & Engineering, IMS Engineering College, Ghaziabad, UP, 201009, India

Abstract: *Long term global warming prediction has major importance in various sectors like climate related studies, agricultural, energy, medical and many more. This paper evaluates the performance of several Machine Learning algorithm (Linear Regression, Support Vector Regression (SVR), lasso, ElasticNet) in problem of annual global warming prediction, from previous measured values. The first challenge dwells on creating a reliable, efficient statistical reliable data model on large data set and accurately capture relationship between average annual temperature and potential factors such as concentration of carbon dioxide, methane, nitrous oxide, Sulphur hexafluoride. The data is predicted and forecasted by linear regression because it is obtaining the highest accuracy for greenhouse gases and temperature among all the technologies which can be used. It was also found that CO2 is the plays the role of major contributor temperature change, followed by CH4, then by N20, then by SF6. After seeing the analysed and predicted data of the greenhouse gases and temperature, the global warming can be reduced comparatively within few years. The reduction of global temperature can help the whole world because not only human but also different animals are suffering from the global temperature.*

Keywords: Climate Change, Machine Learning, Greenhouse Gases, global warming

1. Introduction

Global Warming is the rise in the average for long-term temperature of the Earth's climate system. This climate change is occurred due to some physical factors like CO2, N2O, CH4, SF6 etc. These factors are increasingly simultaneously in the weather. For this, the temperature of Earth"s temperature is also increasing. The temperature for this reason was increasing since 1990 caused by emission of greenhouse gases from the modern industry, vehicle, burning of fossil fuels etc. The average temperature of earth is 1 degree Celsius higher than 100 years ago. Now, many scientists are expecting that within the next 200 years the average temperature will increase by approx. 6 degree Celsius than recent temperature. Global Warming occurs when the greenhouse gases absorb sunlight and solar radiation has reflected from the earth"s surface. For that the atmosphere of the earth is increasing and ultimately the overall temperature of earth is also increasing. The result of the global warming is lethal.

For the global warming the glacier is melting day by day. For that the average height of the sea level is also increasing and this is cause for flooding in many coastal areas. Not only that, but also the underground water of the earth is decreasing. The ocean's temperature is also increasing because of this reason. For that many storms are taking place which has fatal result. Many diseases like allergies, asthma and many other skin problems is also facing by the human for his reason. Many scientists are thinking that, if the temperature will increase like this then many cities will be go under the ocean or submerged with the ocean. A paper [6] describes about global warming. They explain the causes, effects and probable solutions for global warming. There is a paper [7], which gives us some idea about the temperature prediction. This paper helps us to predict the air temperature, which is the main problem of global warming. "Prediction of the likely impact of climate change on monthly mean maximum and minimum temperature in the Chaliyar river basin, India, using ANN-based models" [8], "Big Data and Climate Changes" [9] and "Atmospheric Temperature Prediction using Support Vector Machines Algorithm"[10] are some reference papers of temperature, air condition etc. analysis and prediction. This paper also provided some basic information about temperature, global warming etc. To get relief from the destructive result of Global Warming, many people are reducing the fossil fuel burning to reduce the Global Warming. Not only that but also, they are trying to aware the country people to emit less greenhouse gases. Due to the sea level rise many cities like London, New York, Virginia, Sydney, Charleston, Mumbai etc and sea ports are affected heavily. Due to Global Warming many forests are reducing, many trees are dying for the overheated temperature. For that the ratio of every gases is not perfect in atmosphere in many places. Though the ratio of this effect is lesser comparatively than other. For the high temperature the crops cannot be grown properly which affects the animal food cycle. That means we cannot explain the effects of Global Warming in a single sentence.

2. Literature Survey

There are so many ideas and projects about weather prediction, rainfall prediction, temperature prediction. Some of those ideas are taken for the reference purpose.

In this project the temperature and the greenhouses gases are predicting for next few years. For that the weather prediction-based ideas and projects are focused mainly.

"A new approach for simulating, visualizing and forecasting the rainfall-runoff process within the next two months" by M.J Alizadeh, M.R. Kavianpour, Ozgur Kisi, Vahid Nourani [1] is one of the good concepts for predicting the rainfall. In this ideas Artificial Neural Network, SVR etc techniques are used. This concept is focussing only rainfall. It doesn't provide any information about temperature or greenhouse gases.

"Monthly prediction of air temperature in Australia and New Zealand with machine learning models" by S. Salcedo-Sanz, R. C. Deo, L. Carro-Calvo, B. Saavedra Moreno [2] is also a prediction-based article. In this idea the temperature is only focused. The physical factors are not focused which are responsible for the Global Warming. In this project SVR and multilayer-perceptron methods are used.

"Multiple regression and Artificial Neural Network for longterm rainfall forecasting using large scale climate modes" by F. Mekanik, M.A. Imteaz, S. Gato-Trinidad, A. Elmahdi [3] is also another prediction-based idea. In this project regression and artificial neural network are used to predict the rainfall. This idea is focussing on only the rainfall. It is not focussing on temperature nor greenhouse gases.

"Development and Analysis of ANN Models for Rainfall Prediction by Using Time-Series Data" by Neelam Mishra, Hemant Kumar Soni, Sanjiv Sharma, AK Upadhyay [4] is also used as a reference. In this project regression, mean square error and MRE are used. This idea also focused only on rainfall but not on temperature nor greenhouse gases.

"Application of Artificial Neural Networks to Rainfall Forecasting in Queensland, Australia" by John Abbot and Jennifer Marohasy [5] has also been taken for reference. In this project Artificial Neural Networks is used to observe and forecast rainfall. This idea also doesn't give any explanation about global warming.

3. Proposed System

The main objective of this study is to analyse global warming (temperature and greenhouse gases like- CO2, N2O, CH4, SF6) for approx. 100 – 150 years on the basis data recorded at different place like esrl.noaa.gov, sciencebase.gov etc. The specific objectives are:
1) To predict the temperature and greenhouse gases concentration for next 10 years.
2) To make a graphical interface based on the prediction to make it easy understand.

Machine Algorithms:
There are so many technologies to predict data like svm, linear regression, lasso, elastic net etc. We have tried many of the algorithm to get the highest accuracy. All the methods have different working procedure. The working procedure of some of those predicting techniques are discussed below:

1) **Linear Regression**
   Linear Regression is a method which gives a relationship between a dependent variable or scalar variable and an independent variable or explanatory variable. In this method the relationships are modelled using linear predictor function. Here the data is trained by this method. Linear predictor function is used to make an object of that function and used it for prediction. After creation of the object, the data is forecasted for future.

2) **Support Vector Regression**
   Support Vector Regression is one part of Support Vector Machine. SVR follows the same principle which is followed by SVM. For support vector regression, the prediction method is difficult comparative to other methods. The algorithm is more complicated. Among all the technologies more accuracy is observed in Linear Regression. For that in this paper Linear Regression is used. The complexity on Linear Regression is also comparatively much lesser than the other technologies.

3) **Lasso Regression**

Lasso regression is a kind of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, just like the mean. The lasso procedure encourages simple and sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models that are showing high levels of muticollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

4) **Elastic Net**
   Elastic Net first emerged as a result of critique on lasso, whose variable selection can be too dependent on data and thus unstable. The solution is to combine the penalties of ridge regression and lasso to get the best of both worlds. Elastic Net aims at minimizing the following loss function:

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{m}|\hat{\beta}_j|\right),$$

where α *(alpha)* is the mixing parameter between ridge ($\alpha = 0$) and lasso ($\alpha = 1$).

There are some modules which are required to develop the Global Warming Prediction System. Those modules are briefly explained below:

a) **Data Collection**
   In this module the raw is collected data from different data set. Then the data set is changed as per need. This raw data cannot be predicted directly. So, it is needed to clean and pre-process.

b) **Data Pre-processing**
   In this module the data is cleaned. After cleaning of the data, the data is grouped as per requirement. This grouping of data is known as data clustering. Then check if there is any missing value in the data set or not. It there is some missing value then changes it by any default value. After that if any data need to change its format, it is done. Total process before the prediction is known is data pre-processing. After that the data is used for the prediction and forecasting step.

c) **Data Prediction and forecasting:**
   In this step, the pre-processed data is taken for the prediction. This prediction can be done in any process which are mentioned above. But the Linear Regression algorithm scores more prediction accuracy than the other algorithm. So, in this project the linear regression method is used for the prediction. For that, the pre-processed data is splitted for the train and test purpose. Then a predictive object is created to predict the test value which is trained by the trained value. Then the object is used to forecast data for next few years.

d) **Visualization:**
   In this step, the predicted and forecasted data is used to provide a graphical interface separately. At first the predicted data is plotted in a graph separately with the help of matplot library. Then the forecasted data of temperature is plotted in graph with proper scale. Then the greenhouse gases forecasted data are plotted in a single graph with proper scale.
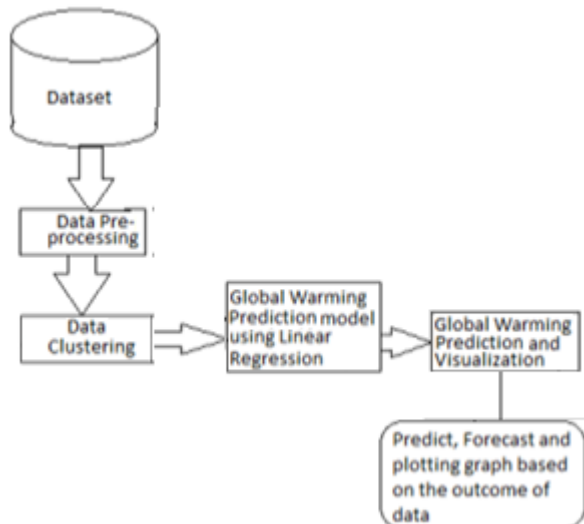
4. Proposed System Architecture

Figure 1: Proposed System Architecture

Dataset Used

a) Temperature

The temperature data set from „sciencebase.gov" is taken. After taking the data, it is modified and reduced as per need. In the raw data, the temperature of each month. Those data are converted into average temperature for better and ease usage. In the temperature data the mean temperature is given. So that we can easily predict the mean temperature by trained those data.

b) Carbon di-oxide:

The data set for the carbon di-oxide is not directly taken data. At first, the data from esrl.noaa.gov" is taken and then the data set is modified as per need. In the raw data set, there are only 20 years data. After that, by referring some other website the data are entered for 70 to 80 years. Then the data is used for further process.

c) Methane

For the methane data set, "esrl.noaa.gov" is referred. In the raw data set of methane, the 80 percent of data is given. Remaining data are filled by referring some other sources.

d) Nitrous oxide

For the N2O data set, „esrl.noaa.gov" is referred. In the raw data set of nitrous oxide, the 90 percent of data is given. Remaining data are filled by referring some other sources. After analysing all the data, 90% data is selected to train. After training the data, an object is created for the linear model. Then, remaining data is predicted and checked the prediction score with the help of that object. Then the temperature and greenhouse gases data are forecasted for the next ten years

e) Sulphur Hexafluoride

For the Sulphur Hexafluoride data set, „esrl.noaa.gov" is referred. In the raw data set of Sulphur Hexafluoride, the 100 percent of data is given.

5. Result and Discussion

5.1 Accuracy Comparison

For CO2 Emission Prediction

Algorithm	R2 Score	Root Mean Squared Error
Linear Regression	0.9852	2.5556
SVR	0.9642	2.5654
Lasso	0.8687	4.8793
Elastic Net	0.9043	2.9747

For CH4 Emission Prediction

Algorithm	R2 Score	Root Mean Squared Error
Linear Regression	0.9366	14.2936
SVR	0.7956	20.3249
Lasso	0.8023	34.8752
Elastic Net	0.7742	31.6324

For N2O Emission Prediction

Algorithm	R2 Score	Root Mean Squared Error
Linear Regression	0.9943	0.3666
SVR	0.9147	1.6722
Lasso	0.9667	0.4368
Elastic Net	0.8931	2.6712

For SF6 Emission Prediction

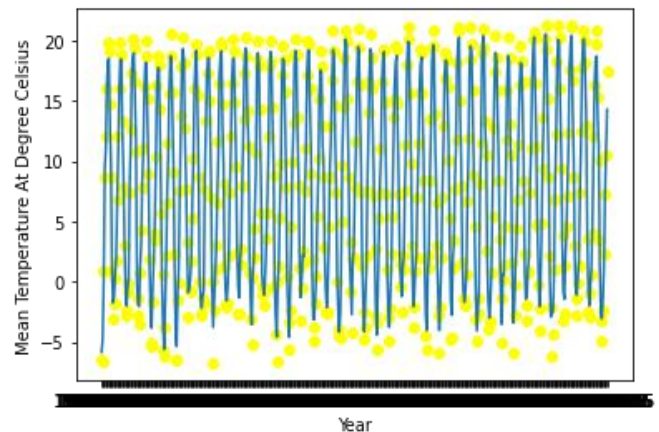| Algorithm	R2 Score	Root Mean Squared Error
Linear Regression	0.9912	0.1680
SVR	0.7545	4.5872
Lasso	0.9276	1.6712
Elastic Net	0.9012	0.9632

Graphical Output

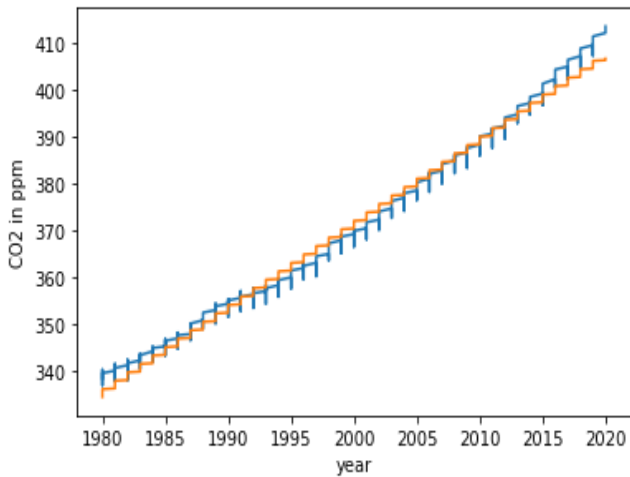**Figure 2:** Year (1980-2020) vs Global Temperature Prediction
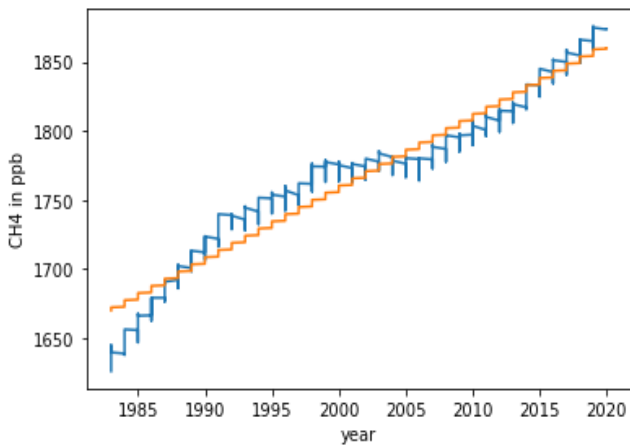
Figure 3: Year vs CO2 Prediction

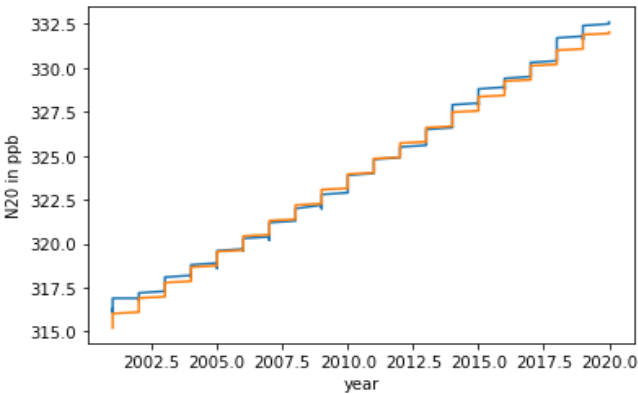
Figure 4: Year vs CH4 Prediction

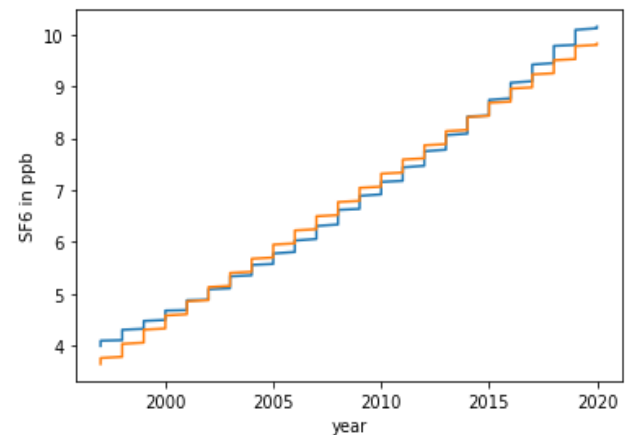**Figure 5:** Year vs N2O Prediction
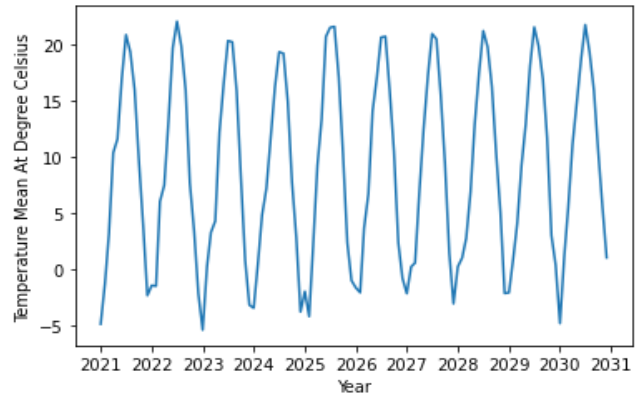

Figure 6: Year vs SF6 Prediction

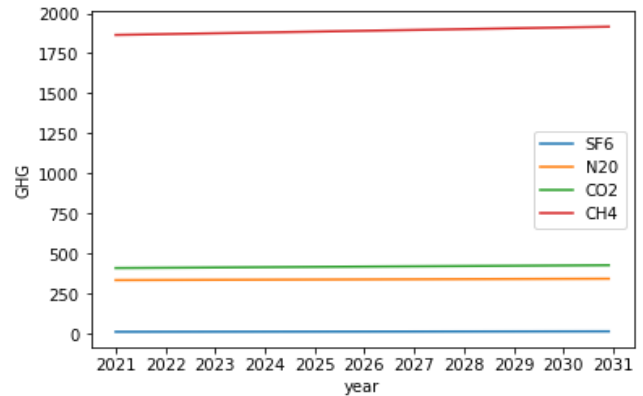**Figure 7:** Year vs Global Temperature Forecasting

Figure 8: Year vs GHG Forecasting

After successful training of the temperature and the greenhouse gases data, more than satisfactory accuracy is achieved for the temperature prediction as well as the greenhouse gases prediction. After the successful prediction, forecasted data for the next 10 years and also graphical representation of those predicted and forecasted data with the help of matplot library is obtained. The graph (Fig: 2) is plotted with the help of the predicted data of year versus temperature. this graph shows the actual mean global temperature that is recorded since 1980 and the predicted value that was calculated during the training of the regression model. The graph (Fig: 3) is plotted between the predicted carbon dioxide and year. The carbon dioxide emission was measured in part per million. This record contains the rise in concentration of carbon dioxide over several decades. The next graph (Fig: 4) is plotted between predicted data of methane and year. Which graph shows the actual rise in concentration of methane since 1985 and predicted rise in concentration of methane over the years. Then graph (Fig: 5) is plotted between nitrous oxide and year. This graph shows the rise in concentration of nitrous oxide since 2002 and the predicted rise in concentration of nitrous oxide. The Last prediction graph (Fig. 6) is plotted between Sulphus Hexafluoride and Year. This graph shows the rise in concentration of sulphur hexafluoride since 2000 and the predicted rise in concentration of SF6. For the forecasting part, the temperature is plotted is separately (Fig: 7) because the unit is different for the temperature from the others physical factors. This graph shows forecasting of global mean temperature 10 years that is from 2021 to 2031. And for the CO2, N2O, SF6 and CH4 are plotted in a same graph (Fig: 8) where the data are compared between years and ppb or ppm (for only CO2). This graph basically shows

forecasting of rise in concentration of different greenhouse gases that is carbon dioxide, nitric oxide, Sulphur hexafluoride and methane in next 10 years that is from 2021 to 2031. Because of using Linear Regression methodology, the graph is observed in a linear pattern (i.e.: $Y = MX + C$), where the X axis represents the years every time whereas the Y axis represents the temperature and other physical factors in different unit. After seeing the graph for the forecasting, it can be said that the temperature and physical factors which are responsible for global warming will increase day by day in a linear pattern and the result of this global warming will become more and more lethal. And soon there will be a need of a solution for this problem.

6. Conclusion

In this paper, the data (temperature and greenhouse gases) of 100-150 years is analysed. Linear Regression and Linear model are used to predict and forecast the temperature and greenhouse gases for the next 10 years in average. The matplot library is used to plot the predicted and the forecasted data. So, at last the following conclusion can be drawn – A model for forecasting data for next 10 years is trained and tested with different input variables like temperature, carbon di-oxide, methane, nitrous oxide, sulphur hexafluoride by linear regression. Some graphs are plotted as a graphical interface for the predicted and forecasted data for all the inputs with the help of matplot library

7. Limitation and Future Work

There is some limitation in this study or this project. This model is predicted only the mean temperature and only 4 green house gasses concentration. It doesn't't explain each week's and each country's data separately. The data set is used for this model for 100 – 150 years. But the prediction will be better if we go more than 200- or 250years data. In future the prediction can be better for using a greater number of data. The other physical entities which are responsible for temperature increasing can be also predicted. The data can be predicted for each week as well as each country, which will make this more attractive and reasonable for all people.

Acknowledgement

References

[1] "A new approach for simulating and forecasting the rainfall-runoff process within the next two months" by M.J Alizadeh, M.R. Kavianpour, Ozgur Kisi; VahidNourani, Journal of Hydrology (Vol – 548); May-2017

[2] "Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms" by S. Salcedo-Sanz, R. C. Deo, L. Carro-Calvo, B. SaavedraMoreno; Theoretical and Applied Climatology (Vol – 125); July-2016

[3] "Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes" by F. Mekanik, M.A. Imteaz, S. Gato-Trinidad, A. Elmahdi; Journal of Hydrology (Vol – 503); October-2013

[4] "Development and Analysis of ANN Models for Rainfall Prediction by Using Time-Series Data" by Neelam Mishra, Hemant Kumar Soni, Sanjiv Sharma, AK Upadhyay; International Journal of Intelligent Systems and Applications; January-2018

[5] "Application of Artificial Neural Networks to Rainfall Forecasting in Queensland, Australia" by John Abbot and Jennifer Marohasy, Advances in Atmospheric Sciences; July-2012

[6] "Global Warming: Causes, Effects and Solutions" by UmairShahzad; Durreesamin Journal (Vol – 1), August2015

[7] "Support vector regression with reduced training sets for air temperature prediction a comparison with artificial neural networks" by Robert F. C, GerritHoogenboom, Ronald W. McClendon, J. A. Paz; Neural Computing and Applications (Vol – 20); February-2011

[8] "Prediction of the likely impact of climate change on monthly mean maximum and minimum temperature in the Chaliyar river basin, India, using ANN- based models" by N.R. Chithra, S.G. Thampi, S. Surapaneni, R. Nannapaneni, A.A.K. Reddy, J.D. Kumar; Theoretical and Applied Climatology (Vol – 121); August-2014

[9] "Big Data and Climate Change" by H. Hassani, Xu Huang and E. Silva; February-2019

[10] "Atmospheric Temperature Prediction using Support Vector Machines" by Y.Radhika and M.Shashi; International Journal of Computer Theory and Engineering (Vol – 1); April-2009

Author Profile

Himanshu Vishwakarma is a B. Tech 3rd year Student in department of Computer Science & Engineering College, Ghaziabad, UP, India. He has done data analyst internship from EdulytInda with successful work on 4 different projects which are Website Blocker, Interactive Dictionary, Credit Card Approval Prediction with Machine Learning and Loan Approval With Machine Learning.