Payment Date Prediction Using Machine Learning

[1]Rithika Pullela, [2]Krishnasri Vaddepally

[1]Information Technology,
[1]BV Raju Institute of Technology, City, Narsapur, India

***Abstract*—The end users of "payment date prediction" are banks and customers of banks who use credit cards. This application extracts earlier paid credit card bills and predicts the date of paying credit card bills for upcoming months. This helps the bankers to generate money consequently.**

Sometimes customers ignore or neglect to pay their credit card bills. This can be overcome by the feature reminders. The customers get reminders every month on a specific date, as the bank analyzes the date of payments of previous months of customers. If customers pay their credit card bills on time the bank can allow for an increase in credit card limit. If customers do not pay bills on time the bank decreases the credit card limit. As the date of payments is being analyzed by banks, this helps the bank to bifurcate customers easily and take action accordingly.

***Index Terms*—Feature Engineering, Feature Selection.**

I. INTRODUCTION (HEADING 1)

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data. The field of study known as machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.

1.1 Definition of Machine Learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks T, as measured by P,improves with experience E.

1.2 Types of machine learning problems

1. Supervised learning: The model or algorithm is presented with example inputs and their desired outputs and then finding patterns and connections between the input and the output. The goal is to learn a general rule that maps inputs to outputs. The training process continues until the model achieves the desired level of accuracy on the training data. Some real-life examples are:Image Classification: You train with images/labels. Then in the future you give a new image expecting that the computer will recognize the new object.Market Prediction/Regression: You train the computer with historical market data and ask the computer to predict the new price in the future.

2. Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. It is used for clustering populations in different groups. Unsupervised learning can be a goal in itself (discovering hidden patterns in data).Clustering: You ask the computer to separate similar data into clusters, this is essential in research and science.High Dimension Visualization: Use the computer to help us visualize high dimensional data.Generative Models: After a model captures the probability distribution of your input data, it will be able to generate more data. This can be very useful to make your classifier more robust.

1.3 Supervised Learning : Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross validation process. Supervised learning helps organizations solve for a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox.Supervised learning uses a training set to teach models to yield the desired output.This training dataset includes inputs and correct outputs, which allow the model to learn over time. The algorithm measures its accuracy through the loss function, adjusting until the error has been sufficiently minimized.Supervised learning can be separated into two types of problems when data mining—classification and regression:

- Classification uses an algorithm to accurately assign test data into specific categories. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labeled or defined. Common classification algorithms are linear classifiers, support vector machines (SVM), decision trees, k-nearest neighbor, and random forest, which are described in more detail below.
- Regression is used to understand the relationship between dependent and independent variables. It is commonly used to make projections, such as for sales revenue for a given business. Linear regression, logistic regression, and polynomial regression are popular regression algorithms.

1.4 Objective of Project : To build a predictive model that can be used to help the Banks use their data efficiently to make better decisions. A predictive analytics application allows the banks and other financial institutions to identify the risks and address them in real time to reach better outcomes.Bank must be able to analyze available data related to the customers before making the decision of issuing a credit card. The model developed will use all possible factors and data to predict whether the customer would fail or succeed in making the next payment with a rational accuracy. lt would benefit the bank before they make any decisions against that customer. The target is to minimize the risk of having loan loss.

1.5 Limitations of Project : Our proposed system is only useful to predict the Date of Payment. If there are some invoices which are overdue for more than a few months which are technically in dispute. So, using our prediction algorithm we cannot identify the invoice which is likely to get into any sort of dispute. But this can be enhanced in future.

II. Description of Algorithms

2.1 Linear Regression

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Steps for Algorithm:

Getting started in R
Step 1:Load the data into R
Step 2: Make sure your data meet the assumptions.
Step 3: Perform the linear regression analysis.
Step 4: Check for homoscedasticity.
Step 5: Visualize the results with a graph.
Step 6: Report your results.

2.2 XGBoost

Steps for XGBoost Algorithm:

Step 1: Make an Initial Prediction and Calculate Residuals.
Step 2: Build an XGBoost Tree.
Step 3: Prune the Tree.
Step 4: Calculate the Output Values of Leaves.
Step 5: Make New Predictions.
Step 6: Calculate Residuals Using the New Predictions.

III. Testing And Validation

Testing is done to look for mistakes. Testing is the process of looking for any flaws or weaknesses in a piece of work. It offers a
a method for evaluating the functionality of individual parts, subassemblies, assemblies, and/or a finished product. It is the process of testing software to make sure that it satisfies user expectations and meets requirements without failing in an unacceptable way. Several test types exist. Every test type responds to a certain testing requirement.

The proposed work is implemented in Python 3.6.4 with libraries scikit learn, pandas, matplotlib and other mandatory libraries. The training dataset is date prediction. Machine learning algorithms are applied such as XG Boost, Linear regression and Random forest. We used these machine learning algorithms and identified intrusion. The result shows that XG Boost is efficient using XGBoost. Random forest achieves 97% accuracy, KNN achieves 94% accuracy,Linear regression achieves 95%,XG Boost achieves 99%.

VALIDATION:

To validate a system, one must watch how it behaves. The verification and validation processes serve as a check to make sure that a phase's output is consistent with both its input and the system's overall needs. To make sure it is consistent with the desired output, this is done. If not, use specific mechanisms for repair to satisfy the standards.

IV. Results

The above-conducted experiment results are to be noted down in a table that specifies the accuracy of each algorithm separately:

Table 1 Experimental Analysis Results

Algorithm	Accuracy(%)
Linear Regression	95.00
Boost	99..00
Random Forest Classifier	97.00
KNN	94.00

V. SCREENSHOTS

Fig.1. data cleaning

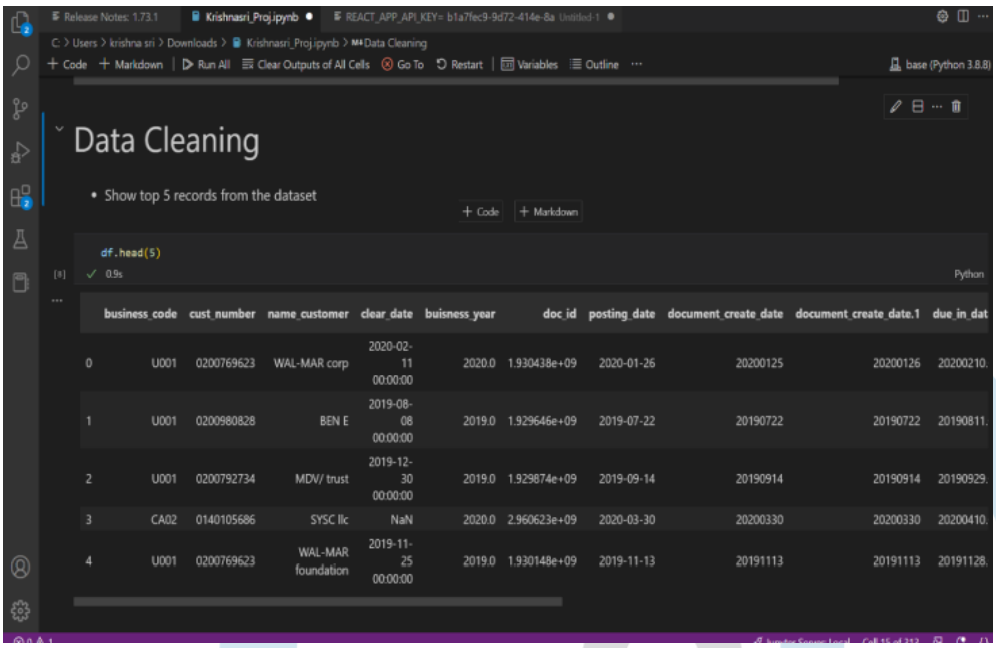


Fig 2. Screenshots of code to drop columns

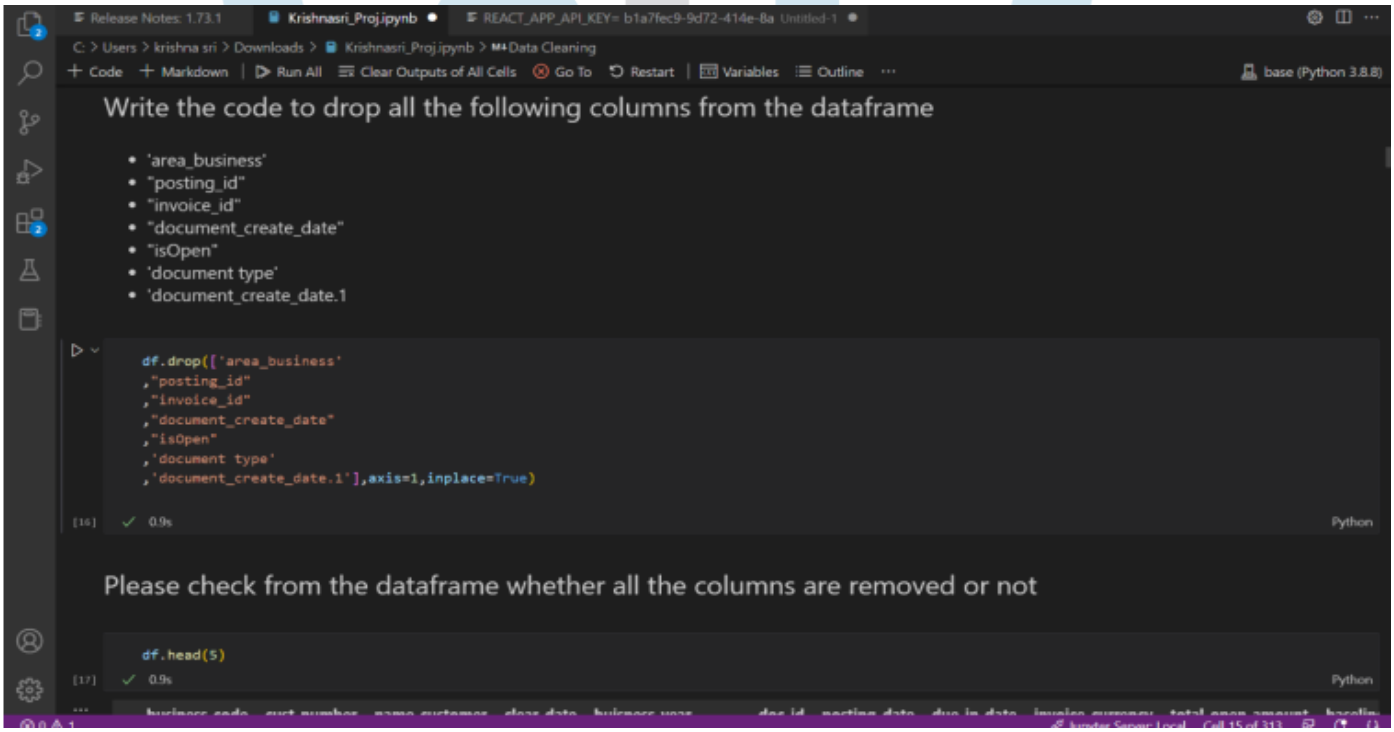


Fig 3: Screenshot of Splitting data into dataset

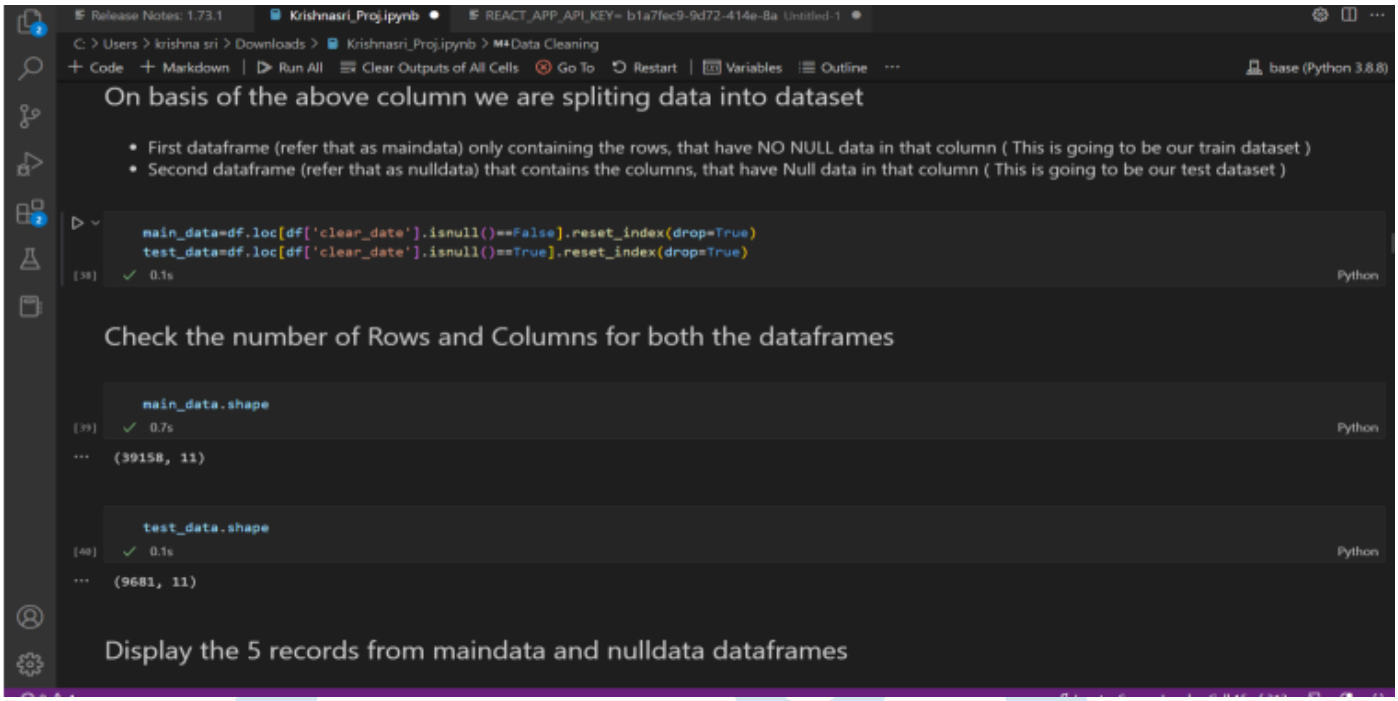
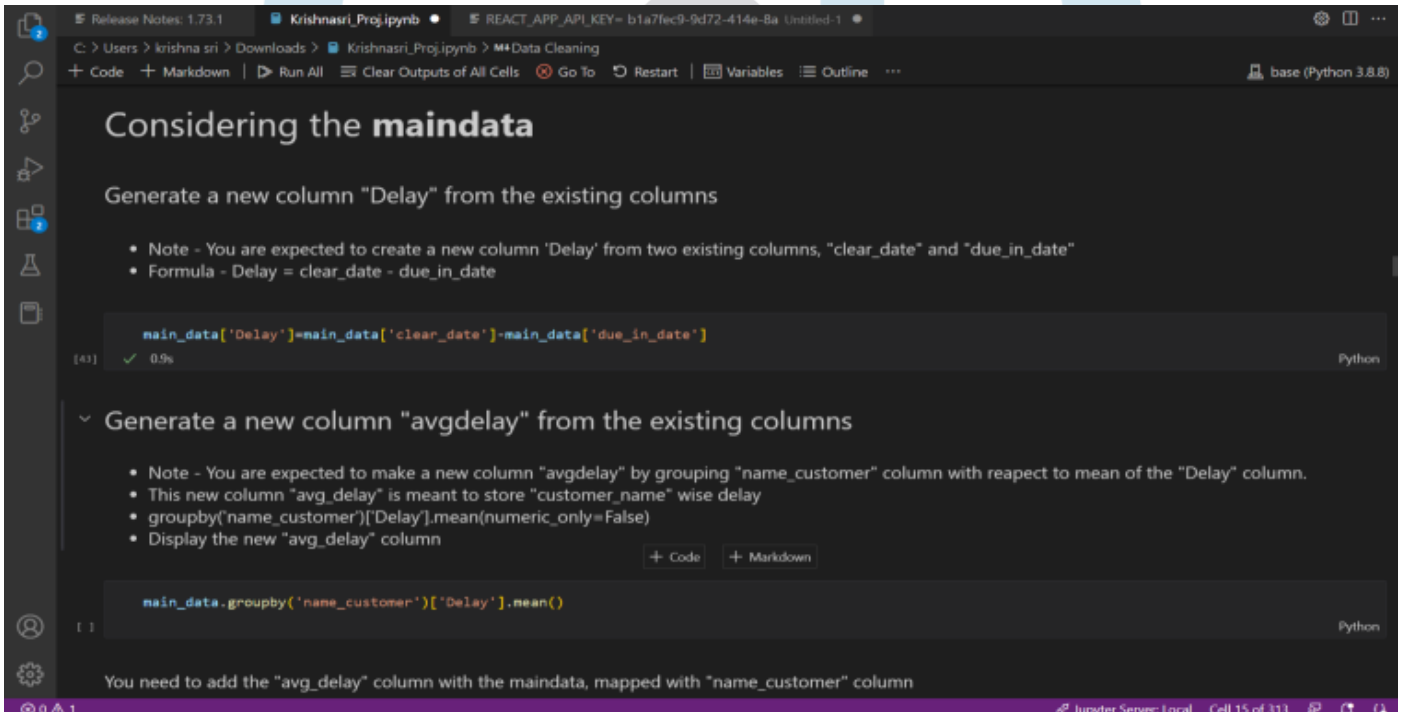
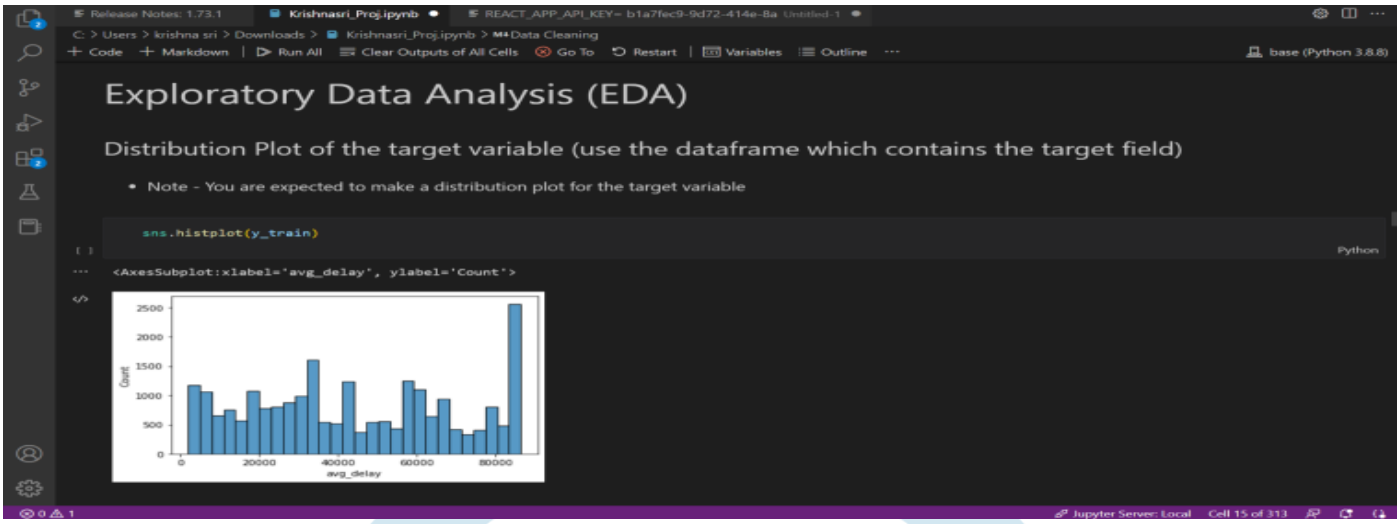


FIG 4: GENERATE DELAY DATE





VI. Conclusion

As the popularity of machine learning is increasing in the finance domain, this is encouraging researchers to identify the problems in the finance domain which can be solved using machine learning techniques. This proposed method can help researchers to get one step closer to achieve full automation in most assets of finance, which is Accounts Receivable. Machine learning is a natural fit to leverage the power of business analytics. Proposed method can help to understand the pattern of invoice and improve the account receivable collection by predicting the delay in advance. This method can also help big business firms to identify and prioritize the delinquent Invoices to save their collection teams time and resources. Proposed method gives more than 85% accuracy regarding invoice's payment outcome. So, it is safe to implement the proposed method in real lif**e.**

Future Work

Our first target is to improve the accuracy for higher age buckets. Implement prescriptive analysis to give a feasible solution for late invoices. There are some invoices which are overdue for more than a few months which are technically in dispute. So, identify the invoice which is likely to get into any sort of dispute.

VII. Acknowledgment

We deem a distant pleasure to acknowledge my indebtedness to the following people who have helped me in completing the investing of our mini project. We express our profound thanks to Dr.K.Lakshmi Prasad, M.Tech, Ph.D., Principal, B. V . Raju Institute of Technology for his cooperation in completing the Mini Project. We express thanks and gratitude to Prof. K. Dasaradh Ramaiah, M.Tech. (Ph.D.), Professor and Head of Department, Department of information Technology and Internal Guide for his encouraging support in carrying out the Mini Project. We thank Project Coordinators Ms. M.Krishna Prasanna, M.Tech, Assistant Professor, Mr. Rakesh , MTech(Ph.D.),Assistant Professor ,for providing us with an excellent project and guiding us in completing our project successfully. We would like to thank all teaching and non-teaching staff of the Department of Information Technology for their support during our course. Finally, we are thankful to our parents and friends for their constant help and moral support. Above all we thank the almighty for giving me strength and courage to accomplish the task.

References

[1] John G. Salek, Accounts Receivable Management Best Practices, ISBN: 978-0- 471-71654-9, July 2005.
[2] John G. Salek, Accounts Receivable Management Best Practices, ISBN: 978-0- 471-71654-9, July 2005. usion_matrix.html
[3]"Descriptive, Predictive, and Prescriptive Analytics Explained" vnunet.com, para. 2, Jan. 31, 2001. [Online]. Available: https://halobi.com/blog/descriptivepredictive-and prescriptive-analytics-explained/