Nick Fuller

C964: Computer Science Capstone

**Part A: Project Proposal for Business Executives**

Letter of Transmittal

9-10-2022

Mr. Doom Slayer

Beeg Beeg Bank Corporation

666 Lazarus Labs Ln

Red Planet, Mars 66666

Mr. Slayer,

We are aware that the Beeg Beeg Bank Corporation requires a solution to its plateauing profit growth. We are also aware that employees are stressed about workload and management is concerned with employees’ ability to correctly approve or deny credit card applications. An excellent solution to both of these problems is to create a machine learning application that can approve or deny credit card applications easily and accurately.

This application will provide many benefits to Beeg Beeg Bank Corporation. The application can be trained using top-quality data from your best agents, ensuring that credit card applications are approved or denied in strict accordance with company policies. The application can also provide information on data trends to stakeholders, proving whether current policies are having the intended effects. The total cost to develop this application would be $150,000.

My expertise in this field is backed up by a Computer Science degree from WGU. Over the years I have proven my ability to adapt, solve problems, and deliver high-quality deliverables to stakeholders. This project would be handled with the utmost care and completed efficiently on time.

I look forward to working with you.

Sincerely,

Nick

**Part A: Project Recommendation**

Problem Summary

The Beeg Beeg Bank Corporation’s profit growth has slowed in recent years. Upper management is skeptical of employees’ ability to approve only worthy candidates who will pay on time. This proposal describes a machine learning application to be built for the Beeg Beeg Bank Corporation that approves or declines credit card applications. It is a stand-alone application built from the ground up employing the use of the Random Forest algorithm. The application will also be able to generate various visual representations based on organized data.

Application Benefits

The solution will implement a Random Forest algorithm that will appropriately approve or deny credit card applications for the Beeg Beeg Bank Corporation and provide data about the approvals to the company. Machine learning will allow the system to be automated and accurately predict which customers should be given a credit card. Other benefits include the speed of delivery to the customer, improved efficiency of employees, and increased adherence to company policies.

Application Description

The application will provide the user with the ability to select various visualizations of the data. The user will also be provided a form to fill out and submit for the customer. The entered information will then be evaluated by the machine learning algorithm and the credit card application will be either approved or denied.

Data Description

The data will be sourced from a public website(kaggle.com). The data will be contained in a CSV file. There are no restrictions on the use of this data. The application will not save or store any entered customer information.

Objectives and Hypothesis

The Random Forest algorithm will accurately prescribe whether any given credit card application is approved or denied a credit card with an accuracy of over 90%. This will be tested against manual approval data from before the application’s creation.

Objectives:

• Create a superb Random Forest model.

• Train the model with good data.

• Test the model thoroughly.

• Focus on delivering a quality product at each step of development.

• Implement technology that produces great visual data for management and stakeholders.

Methodology

The development of the application will follow the SEMMA methodology. This is a proven machine learning development methodology that will provide structure for the project.

• Sample: A sizeable portion of data will be provided for the model.

• Explore: The data will be explored for patterns and abnormalities.

• Modify: The data will be processed into forms readable by the model.

• Model: A model will be created that provides accurate predictions.

• Assess: The reliability of the model’s results will be evaluated.

Funding Requirements

The combined costs of all personnel, hardware, software, and cloud hosting requirements are $150,000.

Data Precautions

There are no precautions for the data that will be used as it is publicly available. Any data entered into the application will not be saved.

Developer’s Expertise

As a Computer Science graduate, I am qualified to create complex applications that solve business problems. Throughout my years of University, I have been prepared to work on large projects and produce quality deliverables. My experiences have also prepared me to be self-sufficient and learn new concepts, as well as solve problems in new or unique ways.

**Part B: Project Proposal**

Problem Statement

The Beeg Beeg Bank Corporation’s profit growth has started to severely decline year over year. Management is afraid that many new credit card applications that are being approved are not being done so following company policy. Employees have also been complaining of increased workload, no doubt due to the lack of automation and the requirement to fill out forms by hand and send them via fax. Not only does this lead to decreased productivity in the workplace, but the time required to send and receive responses from corporate offices also keeps customers waiting and decreases customer satisfaction overall. The Beeg Beeg Bank Corporation requires an application that can automate these processes and provide employees, stakeholders, and most importantly, customers, with quick, effective, and efficient results.

Client Summary

The Beeg Beeg Bank Corporation has been around for decades. They specialize in all things banking. Checking accounts, credit cards, lending, business loans, car loans, and much more. However, being such a long-standing and accomplished institution, the company holds on to many “old school” business practices. The Beeg Beeg Bank Corporation has struggled to keep up with the times and integrate new technologies, including the web.

A machine learning application that approves or denies customers’ credit card applications automatically can help the Beeg Beeg Bank Corporation by reducing manual work done by employees, and decreasing reliance on disconnected systems. The application can ensure that decisions that benefit the business are always made, and take away the stress of making mistakes from employees. Further, the application will decrease the time to serve the customer, therefore increasing customer satisfaction and ensuring brand loyalty in the long term.

Existing System Analysis

The Beeg Beeg Bank Corporation’s current system exists entirely offline. Today, when a customer walks into a branch to apply for a new credit card, an employee sits down with them to fill out paper applications. After the applications are filled out, the employee then must go to the back and fax the pages to the Beeg Beeg Bank Corporation’s corporate offices. It can take hours to receive a response in the form of a fax that is an approval or denial stamp on the application sent initially by the branch employee.

The existing system is not in any way efficient. It takes far too long, involves multiple employees in multiple locations, and worst of all, leaves customers waiting. The proposed solution will fix all of these issues. Credit card approvals can be done right in the branch, instantly, using the proposed application. The machine learning algorithm can decide on the spot, likely with more accuracy than a rushed corporate employee with a stack of new applications piled high. Again, most importantly, the customer is not kept waiting.

The existing system also lacks any data analysis. Filled-out forms are simply filed away in corporate offices. The proposed system can display data visualizations and show trends in the data to stakeholders, which can be used to improve company policies, and in turn, the application.

Data

Data will be sourced from kaggle.com and is available publicly at the following link:

<https://www.kaggle.com/datasets/dansbecker/aer-credit-card-data>

The data is qualitative and nominal. The data is stored as a CSV file

The ‘card’ column is the dependent variable.

The independent variable columns are:

* ‘reports’
* ‘age’
* ‘income’
* ‘share’
* ‘expenditure’
* ‘owner’
* ‘selfemp’
* ‘dependents’
* ‘months’
* ‘majorcards’
* ‘active’

Data will not be collected by the application, however, this function could be added quite easily in a future update. This would allow the system to store new credit card applications and add the data to the visualizations. New visualizations could also be added to allow the data to more closely be inspected in many different ways.

Data anomalies are not evaluated in the application. Adding this function in a future update could help ensure that the machine learning algorithm is trained in the best possible manner, therefore producing better results for both new credit card applications and data visualizations.

Project Methodology

Seeing as we will be aiming to deliver a completed product to the customer, we will be following the first 4 steps of the Agile development methodology.

1. Concept (current)  
   With the creation of this document, we are fulfilling the first step. This document outlines the project and all the details of its creation and delivery.
2. Inception  
   This phase of the methodology is where teams will be created and everything will be prepared for full-scale development. The development plan, including what work will be done within each sprint, will be approved and solidified here
3. Iteration  
   In this phase, sprints will be completed by teams week by week. Weekly activities will include, the development of the application, testing of the application, generation of pre-release versions, and feedback from stakeholders after the pre-release version has been demonstrated.
4. Release  
   During the release phase, final testing will be carried out in production environments and all bugs will be fixed. A polished and high-quality final version of the application will be delivered to the customer.

Project Outcomes

Upon completion of all deliverables, a machine learning application capable of approving or denying new credit card applications will be delivered to the client. This application will also be capable of producing data visualizations that will be useful to upper management. The application will be hosted in the cloud and will be accessible with any internet connection.

Implementation Plan

The machine learning application will be delivered as a completed product. It will be hosted in the cloud and accessible in any PC browser connected to the internet.

The application will first be developed on local systems. After the product has been produced, and before it is uploaded to the cloud, it will be thoroughly tested on local machines.

Testing will be done from the user’s point of view. The quality of the algorithm will be assessed by comparing the input of data to the algorithm to its output and checking for accuracy. The performance of the algorithm will be measured in time to complete requests.

After the application has been thoroughly tested, it will be uploaded to the cloud. Additional testing will be carried out to ensure that the product meets quality standards and is ready for the client’s production environment.

Evaluation Plan

|  |  |
| --- | --- |
| **Objective** | **Success Criteria** |
| Ease of Use | A sample of 10 employees can use the software with 0 issues after a 2-hour training session with a success rate of 90%. |
| User error rate reduction | Previously manually submitted applications are revisited and processed using the new application. We are expecting to see a user error difference of at least 20%. |
| Algorithm Efficiency | The algorithm needs to run in under 5 minutes in a production environment. |

Resources and Costs

|  |  |  |
| --- | --- | --- |
| **Resource** | **Description** | **Cost** |
| Developers | Developers of the code and algorithms | $50,000 |
| UI/UX Designers | Designers of the application GUI and user experience | $40,000 |
| Equipment and other various resources | Computer equipment, software, cloud, etc. | $60,000 |
|  | **Total** | $150,000 |

Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestones** | **Start** | **End** | **Tasks** |
| Week 1 | 9/26/2022 | 9/30/2022 | Random Forest Algorithm Creation |
| Week 2 | 10/3/2022 | 10/7/2022 | UI/UX Creation |
| Week 3 | 10/10/2022 | 10/14/2022 | Visual Data Generation Algorithm Creation |
| Week 4 | 10/17/2022 | 10/21/2022 | Application Finalization (Bring it all together) |
| Week 5 | 10/24/2022 | 10/28/2022 | Testing/Debugging/  Polishing |
| Week 6 | 10/31/2022 | 11/4/2022 | Testing/Debugging/Polishing (cont.) |
| Week 7 | 11/7/2022 | 11/11/2022 | Product finished and submitted for an internal product review |

**Part C: Application**

Link to Binder environment hosting interactive notebook:

https://mybinder.org/v2/gh/nfull14/capstone/HEAD

\*Note about security:

This application does not have any security needs as all data is public, no data is saved, and the business is not protective of this code(it allows it to remain public). In the event customer data needs to be stored in the future, all data would be appropriately encrypted and stored in a password-locked environment implementing least privilege security practices.

**Part D: Post-implementation Report**

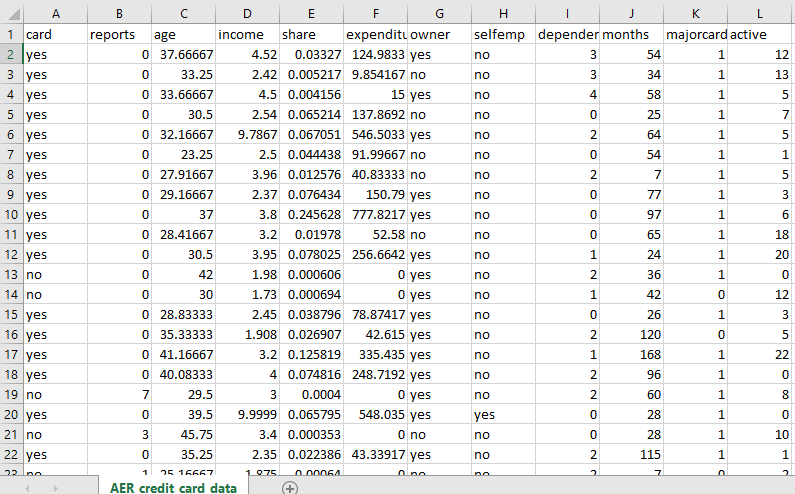
A Business Vision

The Beeg Beeg Bank Corporation required an application that allowed employees to work more efficiently, doing less work by hand, and approved new credit cards more strictly with adherence to company policies. The application also needed the ability to prove its effectiveness by showing biases in the data used to train the model.

The application provided fulfills these needs. The application can produce visuals that show biases in the training data, indicating how the machine learning algorithm will decide to approve or deny new credit card applications. The user is also able to use the application to enter customer information that will be used by the algorithm to approve or deny a new credit card for the customer.

The user fills out the new credit card application for the customer by entering the following information: the number of derogatory reports on the customer’s record, birth month and birth year of the customer, Graphical user interface, text, application, email

Description automatically generatedcustomer’s yearly income, customer’s average monthly credit card expenditure, whether the customer is a homeowner, whether the customer is self-employed, the number of dependents the customer has, the number of months the customer has lived at their current address, the number of major credit cards the customer holds, and the number of active credit accounts held by the customer. After entering values in each of these fields, the user then submits the application and the algorithm immediately returns whether the credit card application is approved or denied.

Datasets

The data, in its raw form, is a collection curated from top employees at the Beeg Beeg Bank Corporation(those that have been known to adhere best to company policies regarding acceptance or denial of new credit cards for customers). The data is stored in a simple CSV file. The data is then processed by the application for the algorithm by simply replacing all ‘yes’ values with ‘1’ and all ‘no’ values with ‘0’. The data is also transformed using a standard scalar before being used by the algorithm.

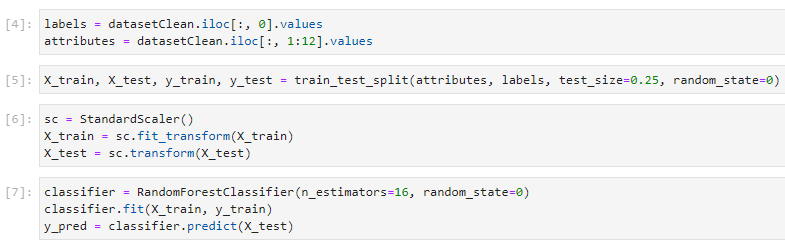
All data used by the application is public and no customer data is stored, therefore there are no security needs for the application in its current form. The application itself is only shared with those involved with the project inside the company on password-protected machines.

Data Product Code

To begin, the CSV file is read and stored in a DataFrame object. A version of the dataset is then produced that is preprocessed for use by the algorithm. This replaces all ‘yes’s and ‘no’s with ‘1’s and ‘0’s.

Three visuals are then produced(using seaborn) using the data and presented to the user, who can use a dropdown box to choose between the three graphics.

The application uses a Random Forrest algorithm to approve or deny new credit card applications. This is an appropriate method as Random Forrest Classifiers are highly accurate.

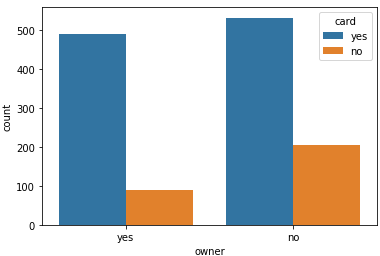


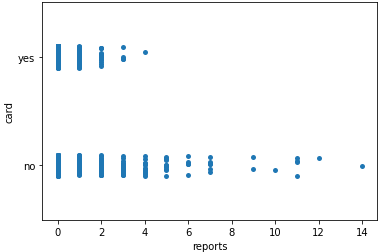
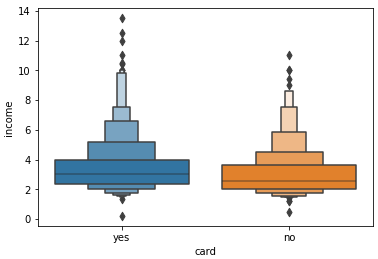
The Random Forrest model used in the application was trained using the data previously loaded and prepared in the application. This data was also scaled before being fitted into the model.

Objective Verification

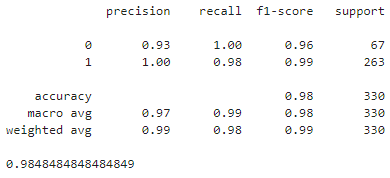
The project’s objective was to produce an application that could approve or deny new credit card applications with an accuracy of over 90%. This objective was met definitively, showing an accuracy of over 96% in most cases.

Effective Visualization and Reporting

While the visualizations produced by the application did not directly influence the development of the used Random Forest algorithm, they are effective in conveying the biases that the trained model will have and what trends its classification decisions are based on.

The application allows the user to explore three different data points about the resulting classification(approval or denial of a new credit card for the customer). The first visual produced is a histogram showing the number of customers approved or denied a new credit card based on their ownership of a home. Analysis of this graphic shows that customers who are homeowners are approved at a higher rate than those that are not. Second, a plot is produced that shows the number of customers approved or denied for a new credit card concerning the number of major derogatory reports on their record. This visualization clearly shows that a higher number of derogatory reports results in the application being denied. Lastly, a boxenplot is produced that compares the number of customers approved or denied a new card based on income. This third visual representation of the data shows there is only a slight bias towards customers with higher incomes being approved for new credit cards at a higher rate.

Accuracy Analysis

The data was split, using 75% for training and 25% for testing. This yielded an accuracy of 98.48% using 16 estimators in the Random Forrest. Changing the number of estimators or the training-to-testing ratio never caused the accuracy of the algorithm to dip below 96% for reasonable values, a great result.

Application Testing

The application was tested as a black box in most instances. Minor tests were conducted using outputs along the way that helped ensure that data handling was done correctly.

After completion of the application, it was tested from a user’s perspective using data from the CSV file to ensure the credit card applications were being approved or denied appropriately.

Future tests, including white box and unit tests, could be developed to increase confidence in the application’s abilities.

Application Files

There are no files required to run the application. Only a modern browser is required to use the application.

Within the Binder application of the browser the file hierarchy is as follows:

/

CreditCardApprover/

AER\_credit\_card\_data.csv

CreditCardApprover.ipynb

README.md

requirements.txt

User Guide

1. Open a modern browser(Chrome preferred).
2. Go to: <https://mybinder.org/v2/gh/nfull14/capstone/HEAD>
3. On the left-hand side, there is a file browser, double-click/open the ‘CreditCardApprover’ folder.
4. Double-click ‘CreditCardApprover.ipynb’ to open the notebook.
5. Click the button that looks like a “fast forward” button on a remote, located between the “restart circle” button on the left, and the download button on the right.
6. If any errors occur(due to Binder, most likely), repeat step 5.

Summation of Learning Experience

My University experiences over the years heavily contributed to my success on this project. Learning multiple programming languages and coding styles has increased my adaptability to new Computer Science concepts. Learning about the power of machine learning algorithms in my previous course(C951) helped me greatly. Applying those concepts to this project proved a fantastic learning experience.

I had to delve deeper into machine learning algorithms and learn more about classification vs regression for this project. After comparing many algorithms, I decided on the Random Forest algorithm due to its high classification accuracy. I also was able to learn a lot about Anaconda and Python, as well as machine learning and graphics libraries like pandas, sklearn, and seaborn.

This project has reinforced the concept of lifelong learning by allowing me to prove to myself that I can seek out my own information and continue to grow and adapt to new projects. This will be invaluable in my professional career. I hope to continue to grow and learn professionally and be part of teams that create amazing products in my future career positions.