WGU C951

Task 3

MACHINE LEARNING PROJECT PROPOSAL

Nicholas Fuller

Student ID #: 001081254

9/2/2022

Table of Contents

A. Project Overview....................................................................................................................pg 3 A.1. Organizational Need ......................................................................................................... pg 3 A.2. Context and Background.................................................................................................... pg 3 A.3. Outside Works Review ......................................................................................................pg 3 A.4. Solution Summary............................................................................................................. pg 4 A.5. Machine Learning Benefits................................................................................................ pg 4 B. Machine Learning Project Design......................................................................................... pg 4 B.1. Scope ................................................................................................................................. pg 4 B.2. Goals, Objectives, and Deliverables ................................................................................. pg 5 B.3. Standard Methodology ..................................................................................................... pg 5 B.4. Projected Timeline ............................................................................................................ pg 6 B.5. Resources and Costs........................................................................................................... pg 7 B.6. Evaluation Criteria ............................................................................................................ pg 7 C. Machine Learning Solution Design....................................................................................... pg 7 C.1. Hypothesis.......................................................................................................................... pg 7 C.2. Selected Algorithm............................................................................................................. pg 7 C.2.a Algorithm Justification..................................................................................................... pg 8 C.2.a.i. Algorithm Advantage ................................................................................................... pg 8 C.2.a.ii. Algorithm Limitation .................................................................................................. pg 8 C.3. Tools and Environment ..................................................................................................... pg 8 C.4. Performance Measurement................................................................................................. pg 8 D. Description of Data Sets ...................................................................................................... pg 8 D.1. Data Source ....................................................................................................................... pg 8 D.2. Data Collection Method..................................................................................................... pg 8 D.2.a.i. Data Collection Method Advantage.............................................................................. pg 9 D.2.a.ii. Data Collection Method Limitation ............................................................................ pg 9 D.3. Data Normalization............................................................................................................ pg 9 D.4. Data Security...................................................................................................................... pg 9 References.................................................................................................................................. pg 9

**A. Project Overview**

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 Forrest algorithm. The application will also be able to generate various visual representations based on organized data.

**A.1. Organizational Need**

An automated system is needed that can approve or decline a credit card application for a customer. This will increase adherence to company policy and reduce human errors in processes. The system also needs to be able to generate visually presented data that shows progress to upper and middle management.

**A.2. Context and Background**

The Beeg Beeg Bank Corporation has seen its profit growth slow in recent years. Upper management is skeptical of employees’ abilities to approve only worthy candidates that will pay on time. Management is convinced that the solution is to use machine learning technology to accurately predict which customers should be approved for a credit card.

**A.3. Outside Works Review**

* Hüllermeier, E. (2021). Prescriptive Machine Learning for Automated Decision Making: Challenges and Opportunities.  
    
  In this article, Hüllermeier outlines the challenges that come with implementing prescriptive machine learning. These challenges include ethical issues and ensuring that the created model is good enough to ensure fairness. The article also suggests the importance of training models using data that is accurate and free from human errors or mistakes. These are all important considerations for this project.
* Peela, H. V., Gupta, T., Rathod, N., Bose, T., & Sharma, N. (2022). Prediction of credit card approval.  
    
  This article provides details of a great implementation of an automated system like the one project will produce. Great insights for our project can be gleaned from this article, including the implementation of machine learning algorithms and applications of produced data.
* Nasiba Mahdi Abdulkareem & Adnan Mohsin Abdulazeez, 2021. "[Machine Learning Classification Based on Radom Forest Algorithm: A Review](https://ideas.repec.org/a/aif/journl/v5y2021i2p128-142.html)  
    
  This article includes great explanations and breakdowns of decision trees and random forest algorithms. This will help ensure that the implementations of these algorithms in the project are done correctly and efficiently.

**A.4. Solution Summary**

The solution is to 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.

**A.5. Machine Learning Benefits**

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.

**B. Machine Learning Project Design**

**B.1. Scope**

In scope:

* Creation of a Random Forest machine learning model.
* Adequate testing to ensure the model behaves correctly.
* Creation of visual representations using the processed data from the model.

Out of scope:

* Any additional functionality not mentioned in this document.

**B.2. Goals, Objectives, and Deliverables**

Goals

• Provide a superb application to the client.

• Support the application for the client for an extended period.

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.

Deliverables

• A fully functional automated system that approves or denies credit card applications employing the Random Forest algorithm that can also display useful visuals for management and stakeholders.

**B.3. Standard Methodology**

Development will follow the SEMMA methodology.

• 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.

**B.4. Projected Timeline**

9/16/2022 – The proposal is accepted by the Beeg Beeg Bank Corporation.

9/22/2022– A technical proof of concept is presented.

11/17/2022 – Submitted for review to the Beeg Beeg Bank Corporation.

12/8/2022 – Deliverables produced.

12/22/2022 – Delivered to the Beeg Beeg Bank Corporation.

**Sprint Schedule**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sprint** | **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 internal product review |

**B.5. 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 |

**B.6. Evaluation Criteria**

|  |  |
| --- | --- |
| **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. |

**C. Machine Learning Solution Design**

**C.1. 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.

**C.2. Selected Algorithm**

The Random Forest algorithm is the selected algorithm.

**C.2.a Algorithm Justification**

The Random Forest is a classification machine learning algorithm that displays a high degree of accuracy. This will be perfect for the binary classification of “approved” or “denied”.

**C.2.a.i. Algorithm Advantage**

The advantages of the Random Tree are the relative ease of implementation and accuracy of the algorithm.

**C.2.a.ii. Algorithm Limitation**

One limitation of the Random Forest algorithm is its slower speed. Computational complexity adds to the lack of speed.

**C.3. Tools and Environment**

Operating System: Windows 10

Programming Language: Python

Libraries: Anaconda

Other tools: Jupyter Notebook

All software used is open source.

**C.4. Performance Measurement**

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.

**D. Description of Data Sets**

**D.1. Data Source**

Data will come from pre-collected CSV files.

**D.2. Data Collection Method**

The data is simply downloaded from a website.

**D.2.a.i. Data Collection Method Advantage**

Data is readily available and easy to use.

**D.2.a.ii. Data Collection Method Limitation**

Data may not be reliable or accurate.

**D.3. Quality and Completeness of Data**

Data will be parsed from files into variables and saved in data structures. This data will be checked for validity and completeness before being handed to the algorithm.

**D.4. Precautions for Sensitive Data**

All stored sensitive data will be encrypted. Machines and programs will be password protected.

**References**

Hüllermeier, E. (2021). Prescriptive Machine Learning for Automated Decision Making: Challenges and Opportunities. arXiv preprint arXiv:2112.08268.

Nasiba Mahdi Abdulkareem & Adnan Mohsin Abdulazeez, 2021. "[Machine Learning Classification Based on Radom Forest Algorithm: A Review](https://ideas.repec.org/a/aif/journl/v5y2021i2p128-142.html)," [International Journal of Science and Business](https://ideas.repec.org/s/aif/journl.html), IJSAB International, vol. 5(2), pages 128-142.

Peela, H. V., Gupta, T., Rathod, N., Bose, T., & Sharma, N. (2022). Prediction of credit card approval. International Journal of Soft Computing and Engineering, 11(2), 1–6. https://doi.org/10.35940/ijsce.b3535.0111222