

PROJECT PROPOSAL

Model for Predicting Credit Card Fraud

Date: October 13, 2023

Prepared by: JJMALK Data Insights, Inc.

Jefford Secondes (C0865112)

Jovi Fez Bartolata (C0869701)

Luz Zapanta (C0879190)

Maricris Resma (C0872252)

Keyvan Amini (C0866360)

Table of Contents

Cover Letter	4
Project Overview.....	6
Approach and Methodology	9
Timeline and Deliverables.....	12
Budget and Pricing	14
Qualifications and Experience	15
Team Members	17
Related Experience	19
Risk Assessment and Mitigation	20
Terms and Conditions	21
Appendix A: Methodology Flowchart	24
Appendix B: Project Schedule Gantt Chart	25
Appendix C: Related Experience	26

Table of Figures

Table 1: Project Timeline and Deliverables	12
Table 2: Project Budget Breakdown	14
Table 3: Project team members role and background	17
Table 4: Risks and Mitigation Plans	20
Table 5: Payment Schedule.....	22
Figure 1: Methodology Flow Chart	24
Figure 2: Project Schedule Gantt Chart	25

Cover Letter

JJMALK Data Insights, Inc.

121 Brunel St.

Mississauga, ON. L2V2V2

analytics@jjmalk.com

(123) 143 4444

William Smith

VP, Credit Card Division

Bank of Mississauga

111 Eglinton Rd.

Toronto, ON. T1V1T1V

smith-william@bankofmissi.com

(416) 153 5454

Dear Mr. Smith:

We are writing to express our genuine interest in working with the Bank of Mississauga on an innovative project to develop a machine learning model for detecting fraud in the bank's credit card transactions that has at least 80% accuracy score. Detecting fraud in credit transactions is a method that enables the bank to offer a wide range of services in the market, which, in turn, increases the customers' trust in the bank's operations. JJMALK Data Insights, Inc. (JJMALK) is motivated by a desire to use advanced data analytics to solve complex problems, and we feel that this proposed collaboration provides a unique potential for both firms to make major progress in improving financial security.

Our commitment to create a cutting-edge fraud detection model designed exclusively for Bank of Mississauga credit cards is outlined in this proposal. We seek to produce a robust solution that not only meets but surpasses industry standards by leveraging our expertise in data science, machine learning, and advanced analytics. The fraud detection model will be delivered in pickle file format to ensure seamless integration in the bank's current

infrastructure. This project's collaborative approach, combining the strengths of JJMALK Data Insights and Bank of Mississauga, promises to produce novel insights and significant advancements in the field of credit card fraud detection.

The key points and areas of specialization in this proposal are based on a data-driven approach that will utilize machine-learning classification algorithms and advanced techniques to extract patterns and anomalies from extensive datasets, establishing the foundation for an effective fraud detection program. To ensure the project's timely completion, collaboration with the Bank of Mississauga is undoubtedly necessary. One of the key components of this proposal revolves around securely analysing credit card transactions. By doing so, we not only mitigate project risks but also enhance the accuracy of the fraud detection model.

In the following letter, you can find the proposal document that explains briefly and accurately how to fulfill this project and clearly detail the schedule, deliverables, budget, and methodology. Below is an overview of our proposal which promises quality delivery that is within the maximum budget of \$90,000 and on-time delivery on December 15, 2023.

Thank you for thinking of JJMALK Data Insights, Inc. as a potential partner in this important project. We are excited about the prospect of collaborating closely with Bank of Mississauga to develop a best-in-class model for detecting fraudulent credit card transactions.

Sincerely,

JJMALK Data Insights, Inc.

Jefford Secondes

Jovi Fez Bartolata

Luz Zapanta

Maricris Resma

Keyvan Amini

Project Overview

Project Title: Model for Predicting Credit Card Fraud

Executive Summary:

The major goal of the project is to develop a predictive modelling application that effectively predicts credit card fraudulent actions, such as identity theft, financial fraud, and unauthorized access. Various supervised machine learning classifier algorithms will be developed and tested upon to be able to identify the model that gives the best accuracy that satisfies the requirement of greater than 80% accuracy. Historical data will be evaluated using various machine learning models, identifying abnormalities and trends suggestive of fraudulent activities. The fraud detection application, aimed to protect the integrity of Bank of Mississauga's operations, shall be used to mitigate fraudulent transactions. This comprehensive approach addresses both known fraud trends and emergent threats, providing a proactive defense mechanism against developing fraud strategies.

Objectives:

- Retrieve and manage credit card transaction Data from Google Cloud Platform
- Create a machine learning classifier model capable of detecting credit card fraud using Logistic Regression Model, Decision Tree, or Random Forest.
- Improve the adaptability of the model to evolving fraud patterns through continuous learning.
- Reduce false positives to guarantee a seamless and trustworthy user experience.
- Deliver the pickle file that can be integrated into Bank of Mississauga's organizational systems and databases to assure seamless operation

Below are the various resources needed in the project:

Database Data Source:

- Google Cloud Platform

Technical Requirements:

- **Programming Language**
 - Python
- **Machine Learning Classification Algorithm**
 - Logistic Regression
 - Decision Tree
 - Random Forest

Technology Requirements:

- **Hardware requirements**
 - High-performance CPU or GPU
 - RAM 128GB more
 - Storage SSD or NVMe
 - Scalability Clustering or distributed computing
 - Security- HSMs or TPMs
 - Data Processing Units
 - Power and Backup
 - Environmental controls
- **Software requirements**
 - Operating system: Windows 10
 - IDE: Visual Studio Code v1.83, Jupyter Notebook
- **Versioning Control Tool**
 - GitHub

To fulfill its goals, the project has numerous critical components. Data collection entails collecting and preparing massive volumes of historical data to ensure its accuracy and usefulness. Feature engineering is used to gain significant insights from data, improving the ability of machine learning models to detect fraudulent behaviors. The foundation of the system is the application of supervised machine learning model which allows for the training and validation of algorithms that can adapt to the dynamic nature of fraudulent actions.

Moreover, this project aims to develop a model that can be easily integrated in the bank's existing infrastructure. This is critical in making the fraud detection system accessible and user-friendly for the bank's fraud analysts.

Approach and Methodology

Classification algorithms are supervised learning techniques used to predict categories based on a set of information. In fraud detection, these algorithms are used to predict if a transaction is fraudulent or not. To build an effective fraud detection system, two models will be developed. The results of the models will be evaluated and the one with better performance will be selected for implementation. The project will be implemented in phases.

1. Data Gathering and Preprocessing

The connection to the Bank of Mississauga's database will be established to ensure that analysts have access to the necessary table/s. This project will use 3 years worth of recent data, and the bank must provide access to cardholders' transactional and demographic information. The bank can provide a separate table specific for this purpose so that only the relevant data, such as transaction date, amount, Merchant Category Code, and merchant location, can be accessed. It must not contain sensitive and Personal Identifiable Information (PII) such as unmasked card number and customer name.

Once database access is established, the raw data will undergo preprocessing to prepare it for analysis. Data profiling and cleansing will be done to ensure that missing data are handled, either by removing them or imputing values. Categorical features will be converted to numeric values using one-hot encoding, and numeric features will be standardized so that the features are comparable and the machine learning models will be able to interpret them on the same scale. Standard Scaler will be utilized to standardize the data and set them to 0 to 1 values.

2. Exploratory Data Analysis

This phase focuses on understanding the data and detecting trends and patterns. The proportion of fraudulent transactions will be computed to assess if there is class imbalance in the data. Analysts will examine the relationship between the target variable and the

independent variables using pairwise correlation to see which features have a strong relationship with the target variable. Descriptive statistics will be utilized to describe the data, and charts will be used to visually summarize information.

3. Data Modeling

Based on the results of data exploration and research, the team will select the appropriate algorithm for the data. In the previous projects handled by JJMALK, the team was able to predict twice by utilizing Logistic Regression and Decision Tree algorithms in classification problems. In this project, the following algorithms will be considered for the data modeling:

- **Logistic Regression:** It is one of the commonly used classification algorithms because of its simplicity and high accuracy. Its ability to set apart observations that belong to binary groups makes it efficient in detecting fraud cases.
- **Decision Tree:** It is a non-parametric approach that uses decision rules that can be visually presented in tree-like structure for easy understanding. It trains in a shorter amount of time but is still reliable in classification problems. K-fold cross-validation will be performed to determine the optimal depth of the decision tree.
- **Random Forest:** It uses a group of decision tree models to classify observations and make predictions. It generally performs better compared to a single decision tree but it takes longer to train.

The models will be trained using 80% of the data, and will be tested using the remaining 20%. The 20% will be used to assess the model's performance when applied to a set of data that it has never seen before.

Class imbalance happens when the proportion of the classes in the data is skewed. It is common in fraud cases because it is vital for companies to manage fraud cases, hence, the number of fraud cases is usually less than 50%. Training the model on data with class imbalance produces unreliable results. To address this, synthetic oversampling will be applied to make the distribution of classes 50-50.

4. Model Evaluation and Final Model Selection

The models will be compared based on their performance in both the training and test data. The difference between the performance of the model in the training and test sets should be within 5 PPS, otherwise, overfitting is a problem. Confusion matrix, accuracy rate, recall rate, F1 score, ROC AUC, and other metrics will be used to evaluate the models. The model with better performance will be selected as the final model for the fraud detection system.

Please refer to [Appendix A: Methodology Flowchart](#) for the flow of the modeling activity.

Timeline and Deliverables

The project consists of 5 Milestones namely Planning, Model 1 Development and Validation, Model 2 Development and Validation, Delivery and Reporting. The project has a total duration of 51 days. Below is a detailed table to show a clear overview on the project schedule with the phases and milestones. Please see [Appendix B: Project Schedule Gantt Chart](#) for a visualization overview on the project timeline.

Table 1: Project Timeline and Deliverables

Project Tasks	Start	End	Deliverables
Planning			
Proposal Creation and Submission	06-Oct-2023	12-Oct-2023	
Data Gathering	13-Oct-2023	20-Oct-2023	
Model Research and Design	13-Oct-2023	20-Oct-2023	
◆ Milestone- Planning Done		20-Oct-2023	Proposal Doc
Model 1 Development and Validation			
GIT Creation	23-Oct-2023	23-Oct-2023	
Data Preprocessing	24-Oct-2023	03-Nov-2023	
Exploratory Data Analysis	06-Nov-2023	10-Nov-2023	
Data Modeling Using ML Algorithm 1	13-Nov-2023	22-Nov-2023	
Model 1 Performance Evaluation	13-Nov-2023	22-Nov-2023	
◆ Milestone - Model 1 Dev and Val Done		23-Nov-2023	
Model 2 Development and Validation (if needed)			
Data Modeling Using ML Algorithm 2	24-Nov-2023	28-Nov-2023	
Model 2 Performance Evaluation	24-Nov-2023	28-Nov-2023	
Model Selection	29-Nov-2023	30-Nov-2023	
◆ Milestone - Model 2 Dev and Val Done		30-Nov-2023	
Delivery			
Code Freeze	01-Dec-2023	01-Dec-2023	
Pickle File Generation and Delivery	04-Dec-2023	04-Dec-2023	

Project Tasks	Start	End	Deliverables
◆ <i>Milestone - Customer Acceptance</i>		04-Dec-2023	<i>zip file of .ipynb and pickle file</i>
Documentation and Reporting			
Documentation	05-Dec-2023	07-Dec-2023	
Create Presentation	05-Dec-2023	07-Dec-2023	
◆ <i>Milestone - Presentation</i>		08-Dec-2023	<i>.doc report and .ppt file</i>
Risk and Event Buffer			
Risk Buffer	11-Dec-2023	13-Dec-2023	
Event Buffer	14-Dec-2023	15-Dec-2023	

Budget and Pricing

The current forecasted cost of this project is \$73,700 which is within the budget requirement of Bank of Mississauga. The table below lists the precise budget breakdown.

Table 2: Project Budget Breakdown

Budget Category	Budget Cost
Software Costs	
Application software	\$500
Operating system	\$1,000
Hardware Costs	
PCs	\$10,000
Network Costs	
Internet	\$1,000
Team Resources	
Project Manager (\$30/hr)	\$12,240
Technical Lead (\$30/hr)	\$12,240
Developer (\$30/hr)	\$12,240
Researcher (\$30/hr)	\$12,240
Validator (\$30/hr)	\$12,240
Total	\$ 73,700

Qualifications and Experience

JJMALK Data Insights Inc., a leading Data Analytics start-up company, is a trustworthy and dependable partner well positioned to provide advanced fraud detection system. With a 5 years track record of success, we have completed various projects that demonstrate our dedication to innovation, data security, and sophisticated analytics.

1. **Advanced Analytics Expertise:** Our team of seasoned data scientists and analysts has extensive experience building and executing advanced analytics solutions. We have a track record of using machine learning algorithms to extract useful insights from large datasets. Previous projects include predictive analytics for credit card customer attrition and transaction anomaly detection systems, which have considerably improved our clients' business operations.

2. **Successful Fraud Protection Solution Implementation:** JJMALK Data Insights Inc. has a track record of successfully deploying fraud protection solutions for a wide range of clients. Our solutions have continually proven their capacity to adapt to changing fraud strategies, resulting in lower fraud rates and increased security for our clients. These projects have spanned a variety of industries, including banking and retail, demonstrating our adaptability in tailoring solutions to suit unique difficulties.

3. **Integration Excellence:** We take pleasure in our ability to integrate systems seamlessly. Our team has effectively integrated complicated systems into existing organizational frameworks, resulting in least disturbance while maximizing productivity. Integrating with various databases, legacy systems, and third-party apps is part of this. Our dedication to integration excellence ensures that the proposed fraud detection system will seamlessly integrate with Bank of Mississauga's existing infrastructure.

4. **Customer happiness and Long-Term Partnerships:** Customer happiness is key to our business strategy. We have formed long-term relationships with clients who have benefited from our solutions, demonstrating our dedication to continued support and collaboration.

Our approach is not simply about delivering a product, but also about ensuring that it evolves in tandem with the needs of your organization.

Finally, JJMALK brings a unique combination of technological competence, industry experience, and a commitment to quality to deliver a cutting-edge fraud detection framework suited to the specific demands of the Bank of Mississauga. We look forward to the opportunity to contribute to your organization's security and prosperity.

Team Members

The team consists of 5 members. See below table for each member's role in the project and corresponding past expertise.

Table 3: Project team members role and background

Name	Role	Background
Maricris Quilop	Project Manager	Maricris is an experienced software developer and project leader who has closely managed software development lifecycles in the IT industry. She has led major development projects and worked closely with other stakeholders in initiating, planning, executing product development, validation, delivery and customer acceptance testing which. She is passionate about data analytics with a strong desire to help with business decision-making for increased productivity and profitability. https://www.linkedin.com/in/maricris-resma-a21090265/
Jovi Bartolata	Technical Lead	Jovi is an experienced data analyst with a demonstrated history working in the banking and financial services industries. She is proficient in various analytical and programming tools including SAS, Python, SQL, and PowerBI. In her previous projects, she developed machine learning models that helped businesses improve acquisition of customers, increase contact rates, and manage customer attrition. https://www.linkedin.com/in/jovi-fez-bartolata/

Jefford Secondes	Developer	<p>Jefford has an extensive expertise in Microsoft SQL Server and other databases like MONGODB and NoSQL, as well as his knowledge of SSIS and SSAS makes him a highly versatile and valuable professional in the field of database management and data analytics. With expertise in NoSQL databases and the broader context of big data, Jefford can work on projects that involve handling vast volumes of unstructured or semi-structured data.</p> <p>https://www.linkedin.com/in/jefford-secondes/</p>
Luz Zapanta	Researcher	<p>Luz is a research fellow at JJMALK Data Insights, Inc., Mississauga Ontario CA. Her current research interest focuses on Machine Learning for healthcare and financial services. She received an M.Sc. in Mathematics from the University of the Philippines. Teaching is one of her passions, and she serves as an instructor and Chair in Math at the Virginia Public Schools, USA.</p> <p>https://www.linkedin.com/in/luz-zapanta-880b6924b/</p>
Keyvan Amini	Validator	<p>Keyvan is a validator to check the deliverables with the customer requirements. He received an M.Sc in IT from the University of Iran. He was an IT consultant in the Banking industry in Iran.</p> <p>https://www.linkedin.com/in/keyvan-amini/</p>

Related Experience

The innovative work of JJMALK inc. in the field of credit card fraud detection has earned the company a reputation as a leader in the big data and analytics industry. Our past project, the "Advanced Fraud Detection and Prevention System", (see [Appendix C: Related Experience](#)) exemplifies our expertise in data modeling, big data analytics, and fraud mitigation. This project's primary objective was to create an advanced fraud detection and prevention system for a major credit card issuer. The client's key objectives were to reduce financial losses due to fraud, enhance customer trust, and maintain compliance with industry regulations.

Our team built a very precise fraud detection model using state-of-the-art data modelling techniques like machine learning, neural networks, and deep learning algorithms. This model was made to sift through mountains of transaction data in search of signs of fraud.

We built a scalable big data infrastructure capable of handling the massive influx of credit card transactions in real-time. This setup guaranteed quick data processing and analysis, which in turn allowed for prompt detection and action against fraudulent activities. Our system used predictive analytics to foresee the possibility of fraud by recognizing nuanced shifts and patterns in cardholder behavior. This foresight assisted in avoiding fraudulent activities.

Risk Assessment and Mitigation

The team has identified 4 important risks and the corresponding action plans to reduce the likelihood of the occurrence of these risks.

Table 4: Risks and Mitigation Plans

Risk	Impact Level	Mitigation Plan
Imbalanced data results in a data model that may lead to bias toward the majority class and poor predictions for the minority class	High	Apply resampling to make up for the imbalanced data
Poor accuracy in data model due to poor quality of data may lead to difficulty that can cause delay	Medium	Apply rigorous data preprocessing to ensure removal of outliers and irrelevant data
The resulting performance metrics of initial modelling algorithms to be used may not reach required accuracy, causing delay in remodelling	Low	Extensive pre-study on the data and research on best model to use must be done beforehand
There is a chance that the project won't be completed in the allotted time due to a team member's illness	Low	The team will apply effective time management by developing efficient work breakdown where there are primary and secondary members assigned to each task

Terms and Conditions

Confidentiality Agreement

All data, information, and intellectual property shared between the parties about the project will be kept secret. A confidentiality agreement will be signed by both sides and each team's members to keep private information safe.

Intellectual Property Rights

JJMALK will own all the software code, algorithms, custom technology, and documentation that are made during the project.

Compliance with Laws and Regulations

The project must follow all laws and rules that deal with finding fraud, protecting data, and any other legal issues that come up.

Termination Clause

A termination clause will be written into the contract. This clause will spell out the circumstances under which either party can end the project. There should be a notice period and a written notice of termination.

Change Management

A change request must be filled out for any changes to the project's goals, schedule, or budget. Changes must be agreed upon by both parties and could affect the project's costs and schedule.

Dispute Resolution

If there is a disagreement or dispute about the project, there will be a formal process for resolving it. As long as both sides agree, this process could include negotiation, mediation, or arbitration.

Warranty

JJMALK promises that the Fraud Detection System will be reliable and efficient. Any problems found during the warranty period within six months after the project is finished will be fixed at no extra cost.

During the project, JJMALK will make sure they have the right insurance, such as professional liability and general liability insurance.

JJMALK agrees to protect and not be responsible for any claims, losses, damages, or liabilities that happen because of the project, including any claims from third parties about the Fraud Detection System.

Payment Terms

Payment terms for the Fraud Detection System project are as follows:

Payment Schedule

Payment will be made in CAD according to the following schedule:

Table 5: Payment Schedule

Milestone	Initial Date	Payment Percentage
Planning Done	20-Oct-2023	20%
Model 1 Dev and Val Done	23-Nov-2023	40%
Model 2 Dev and Val Done	30-Nov-2023	60%
Customer Acceptance	04-Dec-2023	80%
Presentation	08-Dec-2023	100%

Invoicing

Bank of Mississauga shall invoice JJMALK upon successful completion of each project milestone as specified in the payment schedule. Invoices shall be submitted within 2 days of milestone completion.

Payment Due Date

Payments shall be made within [Number of Days] days of the invoice date. Late payments may incur interest charges at a rate of [Interest Rate]% per month.

Retention

JJMALK Insights Inc. may retain a percentage of the total project value as retention. Retention will be released upon successful completion and acceptance of the project.

Payment Method

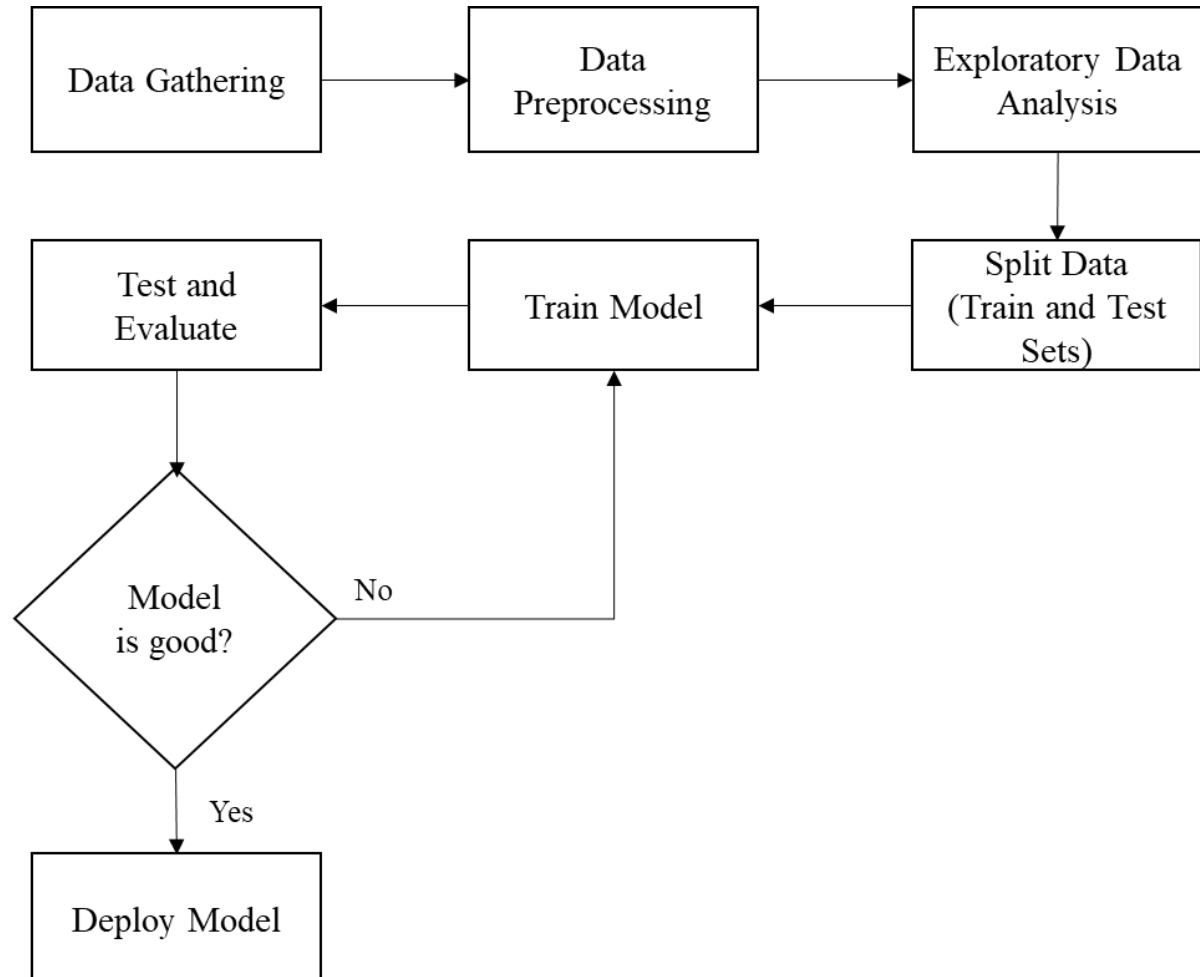
Payments will be made by Cheque to the bank account provided by JJMALK.

Currency

All payments and invoices shall be denominated in CAD and bi-weekly payments

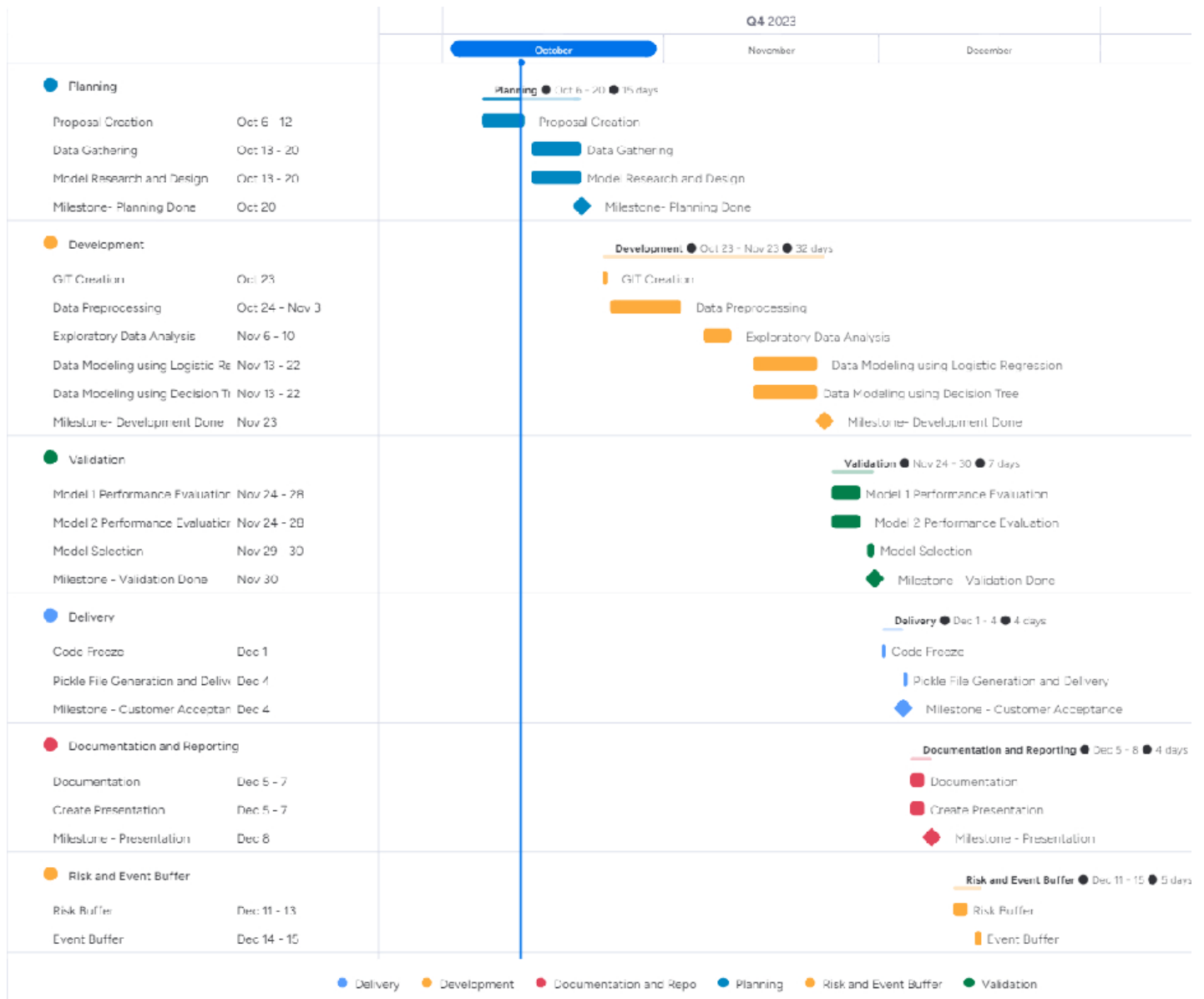
Appendix A: Methodology Flowchart

Figure 1: Methodology Flow Chart



Appendix B: Project Schedule Gantt Chart

Figure 2: Project Schedule Gantt Chart



Appendix C: Related Experience

Sample projects of JJMALK Data Insights, Inc. can be found in:

<https://github.com/JoviBartolata/Projects>