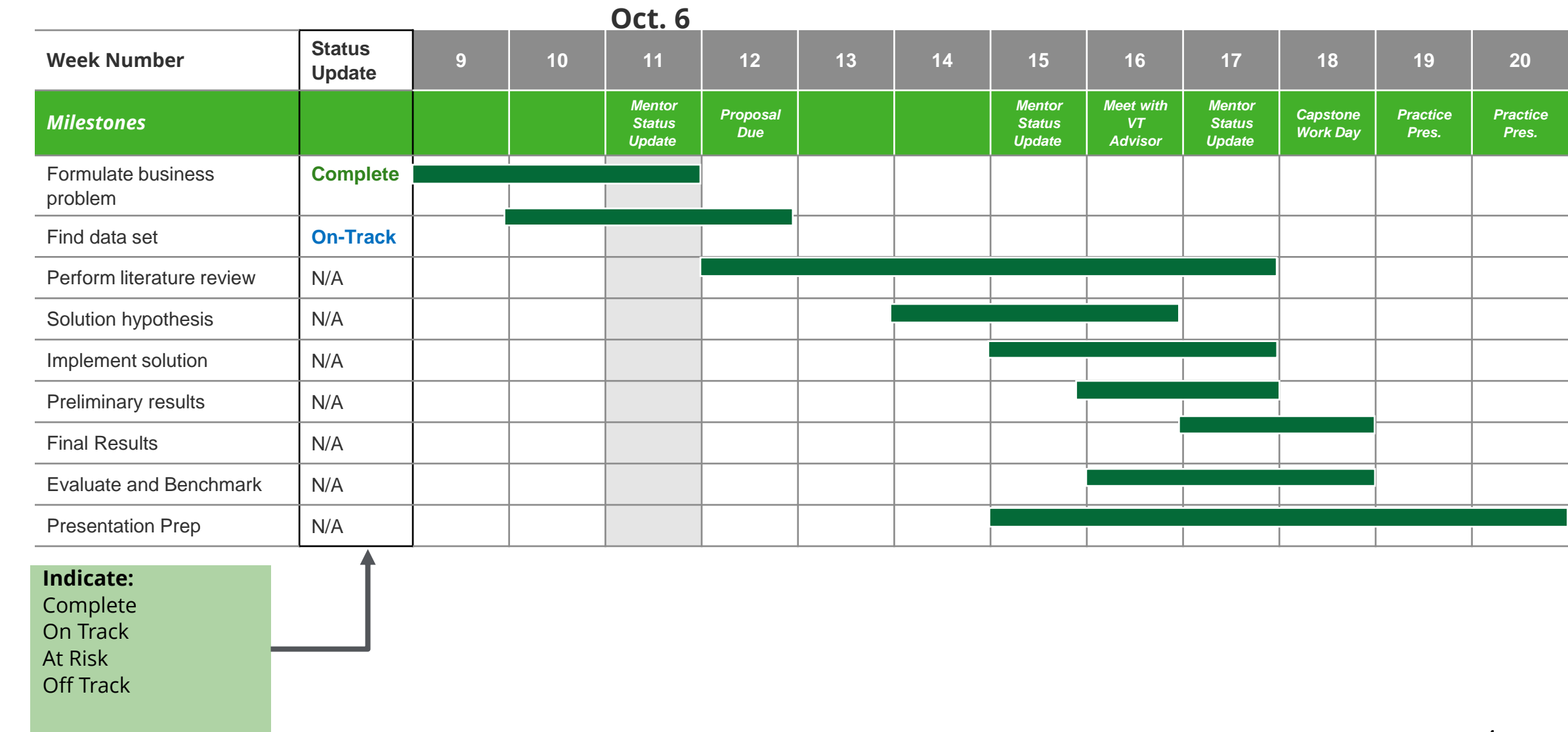


Project Timeline

October 6, 2023



Capstone Proposal: Bank Account Fraud Detection

Project Name		Modeling Requirements	
Business Understanding	Our team is aiming to detect bank account fraud detections by evaluating a variety of variables related to financial status, personal identification, and payment patterns. We will uncover the answers to questions such as: 1. Which factors should be prioritized when monitoring bank account fraud? 2. Which predictors affect fraud detection the most? 3. Are there trends for the frequency of fraud types?	Data Type	CSV file: numeric & categorical variables
		Data Source	Kaggle: Bank Account Fraud
		Data Preparation Steps	Look at distributions, patterns and correlations. Apply appropriate filters for outliers Deal with null values. Standardize variables. Address class imbalance. Train & test split, SMOTE
Client/Project Motivation	Our clients are banks and financial institutions. Bank account fraud costs companies millions of dollars every year and creating a proactive approach towards bank account fraud detection may help cut costs and increase awareness of fraud factors or triggers.	Data Challenges	Target class imbalance, identifying column features such as ambiguous features, limited labeled data
Market /Industry	Finance, Banking, Insurance	Modeling Techniques	Classification: - Logistic Regression - Decision Trees with ensemble methods
State of the Art	Our team plans to make our model state of the art by incorporating logistic regression and building out additional features that other models are currently not considering.	Target Variable	Fraud (1) or no fraud (0)
Success Metrics (Evaluation)	Our team will evaluation success by examining precision, recall, confusion matrix, accuracy rates, ROC and AUC scores.	Regression or Classification problem	Classification
Scalability	The team will build scalability into the model throughout the development stage by experimenting with data set size and varying feature counts to increase the likelihood of continued effectiveness. Employing Logistic Regression as well will improve performance.	Tools/Methodologies	Pandas, SkLearn, NumPy

Status Update

Provide update/reminder on the goal of the project for the client here



What we did in last two weeks



Initial Meeting

Discussed Capstone topic and potential ML techniques to employ



Dataset Research

Researched on Kaggle to source our datasets



Proposal Discussion

Continue to narrow down project details

Brainstorm proposal features



VT Advisor Check-in

Plan for VT Advisor check-ins

What we will accomplish in next two weeks



Finalize Capstone Topic

Decide on a capstone topic and necessary ML techniques and EDA



Meeting Schedule & GitHub

Create a schedule for check-ins, and set up GitHub repo



Begin Proposal Creation (Due Oct. 13)

Begin creating capstone proposal

Roles and Responsibilities

Pengwei Wang: research project topics and datasets

Hadley Campbell: research project topics and datasets

Brian Rodriguez : research project topics and datasets

Payton Stauble: research project topics and datasets