

Data Analytics Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of the following areas so you will have a complete and realistic overview of your project. Your course instructor cannot approve your project topic without this information.

Student Name: Dallas Taylor

Student ID: 001364785

Capstone Project Name: Predicting Loan Approvals Based on Business Rules

Project Topic: Binary Classification Based on Synthetic Features

Research Question: To what extent is a Random Forest Classification model effective at predicting whether an individual will be approved for a loan given realistic synthetically generated loan application information?

Hypothesis: Given realistic synthetic loan application data based on industry standard business rules, a Random Forest Classification model will perform significantly better than a baseline classification model when comparing accuracy, precision, and F-beta scores.

Context: I am a data engineer for a production homebuilder, and the majority of our clients take out mortgages to purchase a home. One of the problems we face is vetting leads for financial qualification. This is not to disqualify homeowners, but does allow our sales staff to recommend options that may be the most beneficial for them. While the loan company would be the official source of qualification or disqualification, it may be useful for us to be able to anticipate a buyer's qualification to have options prepared for them. This project seeks to establish the methodology to accomplish this soft pre-qualification by showing that a Random Forest Classifier can effectively be used to predict loan qualification status.

Data: Realistic Loan Approval Dataset | US & Canada (Kaggle) (Patel 2025)

Data Gathering: The dataset was downloaded as a CSV and will be stored locally.

Data Analytics Tools and Techniques:

Tools: Python (pandas, NumPy, Scikit Learn, matplotlib, Seaborn)

Techniques: After cleaning, I will train a Random Forest Classification model and compare its performance against a baseline classifier using the accuracy, precision, and F-beta metrics. If these metrics are higher than on the baseline model, then this classification model can be viewed as being more effective than always classifying as Approved or Denied, or classifying based on random chance. If the metrics are the same or lower, then this model can be viewed as being less effective as assigning a single classification or classifying based on random chance.



Justification of Tools/Techniques: The reason I am using a Random Forest Classifier is because of its ability to avoid issues related collinearity. This model also tends to guard against overfitting due to its selection process. Finally, this model takes relatively little compute power to perform hyperparameter optimization on, which makes it feasible to train locally instead of on cloud compute resources.

Application Type, if applicable (select one):

☐ Mobile

☐ Web

☒ **Stand-alone**

Programming/Development Language(s), if applicable: Python

Operating System(s)/Platform(s), if applicable: Cross-Platform

Database Management System, if applicable: None

Project Outcomes: This project will produce the following:

- A trained Random Forest Classification Pipeline
- An Exploratory Data Analysis notebook

Projected Project End Date: December 14th, 2025

Sources

Patel, Parth. 2025. *Realistic Loan Approval Dataset | US & Canada*. 11 1. Accessed 12 1, 2025.

<https://www.kaggle.com/datasets/parthpatel2130/realistic-loan-approval-dataset-us-and-canada/data>.

Human Subjects or Proprietary Information

Does your project involve the potential use of human subjects? (Y/N): NO

Does your project involve the potential use of proprietary company information? (Y/N): NO

STUDENT SIGNATURE

_____**Dallas Earl Taylor, 12/1/2025**_____

By signing and submitting this form, you acknowledge that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

TO BE FILLED BY A COURSE INSTRUCTOR



WESTERN GOVERNORS UNIVERSITY

The capstone topic is approved by a course instructor.

COURSE INSTRUCTOR SIGNATURE:



Jim Ashe, Ph.D. Mathematics

COURSE INSTRUCTOR APPROVAL DATE:

Tuesday, December 2, 2025

Project Compliance with IRB (Y/N): Y

