Iteration 2: Application Scorecard

IDMP Group Project

October 15, 2024

1 Project Kickoff

- What are the specific goals of this project?

  Treat the predictor variables and the target variable. Complete pre-processing the data (analysing missing values, missing value treatment as per business logic, outlier treatment, correlation analysis, feature selection, building application scorecard and visualization)

- How do we define the project scope clearly to avoid scope creep?

  The scope entails having an end-to-end solution for generating an application scorecard. Starting with data acquisition, data cleaning, pre-processing, Exploratory Data Analysis(EDA), model building, testing and visualization, we intend to cover the entire gamut of things that are normally done in corporate

- What deliverables must be completed at different phases?
  - Business problem understanding
  - Initial data overview (data understanding)
  - Data quality check
  - Reject Inferencing
  - EDA
  - Feature selection
  - Model Building
  - Visualization report

- What are the major milestones, and what deadlines should we set?

  The above pointers would be the major milestones with 1-1.5 week(s) for every milestone

- Do the team's capabilities align with these goals? Are there any gaps that need to be addressed early on?

  Yes, the team can perform all the above tasks

- Do you have a dataset ready to use for the current project?

  Yes, we are using LendingClub's approved and reject population data Dataset Link

2 Team Discussions

- What are the core skills each team member brings to the table?

  Zankhana – Coding, Visualization and documentation

  Pratiksha – Data Wrangling and documentation

  Rohan – Coding, Data Wrangling

- How will each person's expertise contribute to specific tasks?

  Every TM will work as part of the group after having discussed their plan of action with the group. The work after completion will be proofread and checked by the other 2 TMs thereby making sure that the results are as expected

- What skills are missing that may cause delays or challenges?

  As of now there isn't anything missing as far as skills are concerned. However, a closer look at the data and having an initial QC done remains to be the most critical part of the data pre-processing step. Any unforeseen anomalies in the data could result in challenges/delays

- What tools do we have experience with, and what do we need to learn?

  Currently have experience in - Python, PowerBI, SQL, SAS, Hive and Impala

- What programming languages and platforms should we select based on our project needs and team experience?

  We are going ahead with Python as the primary programming language

3 Skills Tools Assessment

- Are there external resources or team members with expertise in the areas where we lack skills?

  At this point in time, there is no involvement from any external resources. The skill set the team has, seems adequate to resolve any issue(s) that might arise during the project

- Which tools, frameworks, and libraries are most suitable for the project's scope?

  We intend to use the following packages from Python for coding, Pandas, Numpy, Sklearn, Seaborn, Matplotlib etc. The scope of the project would certainly have us include a lot more packages and those will be added as and when required.

- How can we ensure that each team member is comfortable with the tools selected?

  We have had internal group/team discussions and have ensured that everyone in the team is comfortable using the tools selected

- Have specific tasks been assigned based on individual strengths, and are team members clear on their roles?

  Yes, tasks have been decided and distributed based on the skill set of everyone in the team. Having said that, everyone in the team is responsible for educating/sharing the know-how of solving an issue with the other members.

4 Initial Setup

- What development environment setup is necessary for this project?

  Have Anaconda and JupyterLab setup in the system along with PowerBI.

- Have we successfully configured version control (such as Git)? Does everyone have access to the repository?

  Yes, we are intending to use GitHub for version control and maintain a documentation of the same. Everyone in the team would have access to the same repository.

- Have we installed and configured all required software, libraries, and tools?

  Yes, all of us have installed the necessary software on our systems. However, the number and type of packages that we would need to include during the project could vary greatly from what was mentioned earlier.

- What testing can we run to ensure that the development environment is functioning correctly?

  We would have a test sample separate from the development/training environment. This would ensure the robustness of our results.

- What troubleshooting steps should we take if the setup does not work as expected?

  We will have checkpoints setup throughout the code to check if each and every module of the code work as intended

5 Progress Review

- What has been achieved so far? Have we completed the initial setup and repository configuration?

  We have gathered the data but are yet to start working on it. We have created the repositories and GitHub for version control

- Have there been any issues or blockers, and how can we address them quickly?

  There have been no issues/blockers as of now

- Is each team member contributing as expected, and does everyone understand their role?

  As of now, everybody's contribution has been as per expectation.

- Are we on track with the timeline and milestones, or do we need to adjust them?

  Yes, we are very much on track on completing all the milestones as per timeline

- How does the progress align with the project's overall objectives?

  As of now, they are fully aligned to the project's objectives

6 Plan Revision

- Based on progress so far, do we need to adjust the project timeline or milestones?

  Not as of now.

- Are any tasks delayed or requiring reassignment due to workload or skill gaps?

  Not as of now. We have gathered the data and are conducting initial checks on the same to confirm the data quality

- How can we ensure that all members are clear on the revised plan and their next steps?

  If there are changes or revisions to the timeline, regular internal progress meetings would help us ensure that all the members are clear on the revised plan.

- What communication strategies can we implement to avoid future delays or misunderstandings?

  Regular internal meetings and helping TMs to resolve issues/challenges they might face. Above all document all the issues, the method of resolution and the time taken for the resolution during the project. This helps in avoiding future pitfalls

- How will we track progress going forward and maintain alignment with the revised plan?

  Regular internal meetings and checking off the milestones one after the other would help us track our progress going forward

7 Submission for This Iteration

- What specific tasks need to be documented for this iteration's submission?

  For iteration 2 we need a proper project course and timeline for implementation.

- Have we detailed the challenges faced, the solutions implemented, and any adjustments to the plan?

  Initial challenge was finding a dataset that could be the best fit for the business problem we want to solve and also cater to the data management and processing pipelines in order to satisfy the subject needs.

- If your data is available online, please provide a link to access it.

  Yes, here is the dataset we found on kaggle. Dataset Link

- Is the PDF using the Overleaf template, and does it reflect the team's actual progress?

  Yes this PDF is made using the overleaf template.

- Has everyone tracked their progress using the Excel file, and are we submitting it along with the PDF?

  Yes we have kept a weekly track of the project and have submitted with this pdf.

- Does the submission meet all the project requirements, and is it ready for review by stakeholders?

  This is iteration 2 so the project idea has been approved and we are in the implementation phase of this.

- Please provide a link to your GitHub repository with the updated files, including the PDF uploaded there as well. Github link