Iteration 2: Sales Forecasting Project

Ayman Mushtaq Ahmad, Amisha Tiwari, Ronhit Neema

October 15, 2024

Contents

1 Project Kickoff

- The primary goal of this project is to build a robust sales forecasting model using machine learning algorithms to predict future sales trends accurately.
- The scope is defined by focusing on data preprocessing, model comparison, and evaluation without expanding into unrelated domains such as marketing analysis.
- Key deliverables include data preprocessing, exploratory analysis, model implementation, and final performance evaluation at various phases.
- Major milestones include dataset readiness, data preprocessing, model selection, training, and final evaluation, with deadlines.
- The team's expertise in data science aligns with these goals, but we plan to learn advanced ML models like CatBoost and XGBoost which might be needed in this project.
- Yes, the dataset has already been identified but we plan to explore more datasets and choose the best fit for the scope of this project.

2 Team Discussions

2.1 Ayman Mushtaq Ahmad

- **Core Skills**: Ayman has some expertise in data preprocessing, feature engineering, and data analysis using tools like Pandas and NumPy. He Plans to upskill on these in the coming week to ensure successfull implementation of project. His expertise will also help in visualization aspect of the project in later stages.
- **Task Contribution**: Ayman will focus on cleaning the dataset, handling missing values, and performing exploratory data analysis (EDA) to extract meaningful insights. He believes that the required skills have been covered in class already and is confident of implementing it on the chosen dataset.
- **Challenges**: While Ayman is proficient in data handling, expertise in hyperparameter tuning for models like XGBoost and CatBoost is a big challenge, with a potential to slow down optimization. Other team members are planning to learn and help with implementation on this together.
- **Tools Experience**: Ayman is experienced with Jupyter Notebook, Pandas, and Matplotlib but may need to learn advanced tuning tools such as Optuna for hyperparameter optimization. He also plans to learn and support Amisha in ML implementation as it gets covered in the course in future lectures.
- **Languages and Platforms**: Ayman is proficient in Python and will use Jupyter Notebooks for data exploration and visualization, which is well-aligned with the project's requirements.

2.2 Amisha Tiwari

- **Core Skills**: Amisha is planning to work on implementing machine learning models, including decision trees, random forests, and regression models using Scikit-learn and XGBoost.
- **Task Contribution**: Amisha will lead the implementation and evaluation of machine learning models, ensuring that the models are trained and tested effectively.
- **Challenges**: Amisha might need support from Ayman and Ronhit with feature engineering to ensure bottlenecks are not created in preparing the data for more complex models.
- **Tools Experience**: Amisha has experience with Scikit-learn and plans to learn XGBoost, LightGBM and CatBoost as well as she aspires having a career in a similar role in the future.

2.3 Ronhit Neema

- **Core Skills**: Ronhit plans to cover model evaluation, performance metrics, and optimizing models for better accuracy and efficiency. While Ayman and Amisha are working, he plans to take a head start on this aspect of the project.
- **Task Contribution**: Ronhit will be responsible for evaluating the performance of the models using metrics such as R-squared and adjusting models based on feedback from Ayman and Amisha. Ronhit will be working on Tasks with Ayman and Amisha in the Initial phase of the project as well for end to end learning.
- **Challenges**: Ronhit might need more exposure to advanced optimization algorithms, which could delay the model fine-tuning process.
- **Tools Experience**: Ronhit is learning Matplotlib for visualization and evaluation but may need to learn additional tools.
- **Languages and Platforms**: Ronhit is proficient in Python, which aligns well with the project requirements for evaluating and visualizing model performance.

3 Skills & Tools Assessment

- We plan to learn the required skills in lectures and implement the same in our project.
- Scikit-learn, XGBoost, LightGBM, and Matplotlib are most suitable for this forecasting project, offering flexibility in handling structured data and hence all team members are looking to gain skills on these to ensure successful project implementation.
- Team members are familiar with Python and Pandas, but need to learn additional skills to enhance performance.
- Tasks are assigned with the idea that everyone has a clearly defined task while also ensuring people contribute together in each phase of the project for end to end skill development.

4 Initial Setup

- Development environment has been setup.
- Version control with Git is being set up to ensure every team member has access to the project repository on GitHub.

5 Progress Review

- The initial setup is complete, the dataset exploration is in progress with one identified as well but not finalized yet.
- Data Set identified is being evaluated for data quality and checks for missing data.
- Each member is contributing as expected, with specific roles in data cleaning, model selection, and result evaluation while also learning skills needed for the next phases in the project.
- We are on track with our timeline and have charted out next steps and milestones of the project which will be discussed in weekly review meetings.

6 Plan Revision

- Based on current progress, we may need to adjust timelines for model tuning if we face challenges because of limited skill levels on certain elements in the project.
- No major tasks are delayed so far as we are still in the initial phase.
- Clear communication in weekly meetings will ensure all members are aware of any revisions to timelines or deliverables.
- Progress will be tracked through weekly reviews, ensuring alignment with revised goals and immediate feedback on any blockers.

7 Submission for This Iteration

- Tasks for this iteration include documenting the initial setup progress and charting out the plan and timeline for project.
- We are exploring our selected dataset for quality and preprocessing requirement. Challenges like missing data are being addressed using appropriate techniques learned in class, and the plan for implementation is being adjusted accordingly.
- The dataset has been identified but we are still exploring other datasets which could come in handy for our project. Once finalized, we will perform Data Preprocessing steps and upload the same as part of next Iteration.
- The PDF report is prepared using Overleaf and reflects our team's Plan and progress, including ongoing tasks and timelines for the planned activities.
- We have created and shared an Excel file, and it will be submitted as part of next Iteration of the Project.
- The submission is complete and meets the requirements for this iteration, ready for Professor feedback and review.