

Team Contribution Report

Team Number	21 (LLM Models Performance Comparison Analysis)
-------------	---

#	Team Members' Full Name	Contribution	% Of Credit
1	Harish Nandhan Shanmugam	<p>I contributed to the data gathering phase by web scraping information from the Hugging Face LLM Arena Leaderboard. I worked on data cleaning and preprocessing, reducing the dataset from 700k rows to 91k rows. Additionally, I conducted univariate analysis as part of Exploratory Data Analysis (EDA).</p> <p>For model development, I implemented the Random Forest Algorithm and Voting Classifier. I also contributed to creating the key results notebook and drafting the final report, which included documenting the data preprocessing steps, models implemented, key results, and evaluation metrics in a technical paper.</p> <p>Furthermore, I managed the Git version control system, ensuring all updates were properly pushed to the repository.</p>	100
2	Akshara Sri Lakshmipathy	<p>I extensively reviewed research papers and explored various domains to select the problem statement for our project. I suggested Hugging Face as the data source, identifying it as the ideal dataset for developing our idea. I was actively involved in data cleaning, where I addressed categorical columns by applying Label Encoding and One-Hot Encoding techniques. During the Exploratory Data Analysis (EDA) phase, I managed the bivariate analysis plots and developed key performance indicator (KPI) visualizations.</p> <p>In the model development phase, I researched the Light-GBM classifier and successfully implemented it in the notebook. Additionally, I contributed to the technical report by drafting the abstract, introduction, related works, conclusion, and references.</p>	100

3	Shivaraj Senthil Rajan	<p>I took full responsibility for the website development and conducted extensive research to define the objectives and goals of our project. During the data gathering phase, I processed the web-scraped data and converted it into the respective ‘.csv‘ file format.</p> <p>In the Exploratory Data Analysis (EDA) phase, I implemented visualizations such as word cloud segmentation, Kernel Density Histogram plots, and various subplots to analyze different models and languages, which helped uncover significant data patterns.</p> <p>To address the class imbalance issue, I applied the SMOTE technique. In the model development phase, I identified overfitting problems with a standard supervised classification approach. To resolve this, I proposed and implemented ensemble methods, including XGBoost and CatBoost algorithms.</p> <p>For the technical report, I converted the content into a standard ACM format, ensuring consistency and professionalism.</p>	100
---	------------------------	---	-----

Consent

We hereby agree with the above team and contribution information.

#	Team Members' Signatures:
1	Harish Nandhan Shanmugam
2	Akshara Sri Lakshmipathy
3	Shivaraj Senthil Rajan

Date: 12/06/2024
