**SEA 820 Final Project Plan: AI-Generated Text Detection**

**Team Members:** Saloni Kamboj, Apeksha Nanda

**Project Duration:** 3 Weeks

**Week 1: Foundations & Classic Model**

| **Task** | **Assigned To** | **Description** |
| --- | --- | --- |
| GitHub & Colab Setup | Saloni | Create repo, add README.md, invite teammate |
| Data Download + EDA | Saloni | Explore text length, class distribution, vocabulary |
| Preprocessing Pipeline | Apeksha | Implement cleaning, tokenization, and handling length |
| TF-IDF Features | Apeksha | Convert text to numerical format |
| Train Classic Model | Saloni | Train and evaluate baseline with precision, recall, F1 |
| Baseline Evaluation & Plots | Apeksha | |  | | --- | |  |  |  | | --- | | Create visualizations and store metrics for comparison | |
| Finalize Project Plan | Both | finalize the document and work accordingly |

**Deliverable:** TF-IDF + Logistic Regression baseline with metrics.

**Week 2: Transformer Fine-Tuning**

| **Task** | **Assigned To** | **Description** |
| --- | --- | --- |
| Load Dataset | Apeksha | |  | | --- | |  |  |  |  |  | | --- | --- | --- | | |  | | --- | | Load Dataset using Hugging Face Datasets |  |  | | --- | |  | | |
| Tokenization | Apeksha | Truncate/pad input, attention masks |
| Fine-tune Transformer | Saloni | Use Trainer API with learning rate, epochs, batch size tuning |
| Log Experiments & Metrics | Saloni | Record hyperparameters, metrics, confusion matrix |
| Compare w/ Baseline | Apeksha | Use metrics and plot comparison |
| Code Cleanup & GitHub | Saloni | Maintain code and version control |

**Deliverable:** Fine-tuned DistilBERT model, performance comparison.

**Week 3: Error Analysis & Reporting**

| **Task** | **Assigned To** | **Description** |
| --- | --- | --- |
| Error Analysis (FP/FN) | Saloni | Identify frequent misclassifications; analyze false positives/negatives |
| Ethical Implications | Apeksha | Discuss misuse, bias, harm to non-native speakers |
| Draft Final Report | Apeksha | Write Introduction, Methods, Results, Analysis, Ethics |
| Report Review + Visuals | Saloni | Create final report and create presentation and add visuals (any table, chart) |
| Slides & present | Both | Record presentation and submit |

**Deliverables:** Final Report (PDF), GitHub Repo, Presentation Slides.