BDM 3035: BIG DATA ANALYTICS CAPSTONE PROJECT

Submitted to: Meysam Effati

Submitted on: 12/08/2024

SMS SPAM CLASSIFIER

REPORT – 5

GROUP - F

Alwin Kannyakonil Scaria

Anisha Susan Mathew

Ashna Viji Alex

Jobin Philip

Mohamed Afthab

INTRODUCTION

This milestone report provides a comprehensive overview of the SMS Spam Classifier project's latter stages, including advanced model development, model deployment, and documentation. This report explains how the final deliverables were developed, the solution was put into practice, and the machine learning models that were assessed and verified in the previous milestone were improved.

REPORT ON STATUS

Summary of work performed:

1. Model Creation and Implementation:

   - Developed and put into use a trustworthy SMS spam detection system using traditional machine learning models that have been improved for precision and efficacy.
   - Made it simple to integrate with the systems that are in place now, allowing for real-time communication and efficient spam detection.
   - To facilitate efficient and scalable spam detection, real-time prediction APIs were made accessible.

2. Model Execution:

   - High accuracy and minimal latency were guaranteed by deploying the learnt models in real time.
   - Included real-time model performance monitoring, allowing for continual enhancement and required system adjustments.
   - Load testing verified the system's reliability and scalability in a range of traffic situations.

3. Maintaining Documents and Reports:

   - Collected extensive documentation covering all project aspects, including model training, data preparation, deployment processes, assessment measures, and system architecture.
   - Wrote the project's final report, including the stages from inception to implementation as well as the challenges faced, and the solutions discovered

ACHEIVEMENETS AND MILESTONES

1. Effective Model Implementation:

   - The SMS Spam Classifier was implemented in a real-time processing production environment.
   - The system proved resilient to changes in traffic volume, retaining its excellent accuracy and low latency.

2. User-friendly Design of Interfaces:

   - Using HTML, CSS, JavaScript, and Python (Flask), a user-friendly interface was developed so that even a non-techie could review real-time data, monitor model performance, and make modifications.
   - These technologies were used in the development of the interface, which supported more efficient operations and speedier decision-making by making it simpler to access key system indicators and deliver actionable insights.

3. Extensive Recordkeeping and Reporting:

   - Produced a comprehensive documentation bundle that acts as a roadmap for upcoming enhancements as well as a point of reference for stakeholders.
   - The project's goals, methods, outcomes, and future directions were all effectively communicated in the final report and presentation.

CHALLENGES FACED

1. Real-time deployment scalability:

- It was difficult to manage resources and minimize delay while the system was processing a large volume of SMS messages in real time.
- Added dynamic scaling to enable the system to adjust resource allocation in response to traffic needs, guaranteeing effective handling of massive SMS data volumes.
- Improved API endpoints to speed up processing and reduce response times, improving system responsiveness.
- Performed extensive load testing to confirm the system's stability at maximum loads, guaranteeing that it would continue to work and be dependable even in scenarios of high usage.

2. Creating a User-Friendly Interface:

- Developing an intuitive interface that balanced accessibility and utility proved to be challenging, particularly when it came to translating complex model performance indicators into data that non-technical users could comprehend and utilize.
- Added degrees of complexity due to the need for thorough testing and improvement to ensure cross-browser compatibility and responsiveness across several devices.
- Integrating real-time analytics and ensuring interface responsiveness while handling large volumes of data necessitated front-end and back-end component optimizations, which presented another significant challenge.

KEY FINDINGS AND ACCOMPLISHMENTS

- Scalability and Real-Time Processing: While creating scalable real-time systems, this project served as a reminder of the need of load balancing and dynamic resource allocation.
- Complete Documentation: Thorough documentation is essential to the project's longevity and facilitates future modifications and expansions.
- System Integration: Careful planning and optimization are required to integrate complex models into existing systems while maintaining system performance and dependability.

TIMELINE CHARTS

TASK	ORIGINAL TIMELINE	REVISED TIMELINE	STATUS
Data Collection	Week 1	Week 1	Completed
Data Preprocessing	Week 2	Week 2	Completed
Feature Extraction	Week 3	Week 3	Completed
Model Selection	Week 4	Week 4	Completed
Model Training	Week 5	Week 5	Completed
Model Evaluation	Week 6	Week 6	Completed
Advanced Model Development	Week 7	Week 7	Completed
Model Deployment	Week 7	Week 7	Completed
Final Report & Demo	Week 8	Week 8	Upcoming

UPCOMING ACTIONS

- Final Presentation and Submission: Putting together an extensive presentation that highlights the model's performance, system design, and project outcomes.
- System upkeep and monitoring: Putting in place continuous monitoring to make sure the deployed model keeps working effectively and adjusts to any shifts in data patterns.
- Examining Complex Models: looking at the possibility of using sophisticated models like BERT in next upgrades to the system to improve its functionality even further.

CONCLUSION

After hitting many significant benchmarks, the SMS Spam Classifier project has successfully implemented an accurate and efficient spam detection system. The project has demonstrated how to efficiently develop models, manage data in real time, and offer user-friendly interfaces, all backed up by extensive documentation. Novel approaches were utilized to address concerns such as scalability and interface usability, ensuring robust system performance and adaptability. The completion of the presentation, maintenance and tracking of the implemented system, and research into sophisticated model adjustments to further refine and boost functionality will be the main priority moving forward. When all is said and done, this program creates a solid foundation for future expansion and improvement.