

## Project Initialization and Planning Phase

Date	15 March 2024
Team ID	SWTID1727180793
Project Title	Sms-Spam Detection Using NLP
Maximum Marks	3 Marks

### Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	SMS spam detection using NLP aims to automatically classify messages as spam or legitimate, improving user experience and security by filtering unwanted and malicious content. NLP enables efficient, private, on-device detection. Techniques like tokenization and vectorization identify spam patterns for accurate classification.
Scope	The scope of SMS spam detection using NLP includes preprocessing SMS text, extracting spam-related features, and training models to classify messages as spam or legitimate. It involves real-time, on-device detection for privacy and efficiency and requires continuous updates to adapt to new spam patterns. Extending to multilingual support and resource-limited environments is also possible.
Problem Statement	
-	
Description	Create an SMS spam detection system using NLP to classify messages as spam or legitimate by identifying spam indicators like keywords and patterns. It should work efficiently on-device, support multilingual contexts, and adapt to new spam trends.
Impact	This system will improve user experience and security by reducing exposure to unwanted and malicious messages, supporting safer SMS communication, and ensuring privacy through on-device processing.

	Its adaptability and accessibility make it beneficial globally.
<b>Proposed Solution</b>	
Approach	Employing natural language processing (NLP) techniques combined with machine learning to accurately detect and filter spam SMS messages, enhancing security and user experience.
Key Features	<ul style="list-style-type: none"> <li>• Implementation of an NLP-based model to analyze and classify SMS messages as spam or legitimate (ham).</li> <li>• Real-time spam detection on mobile devices, ensuring privacy and efficiency.</li> <li>• Continuous learning and adaptation to evolving spam tactics and multi-language support for broader applicability.</li> </ul>

### Resource Requirements

Resource Type	Description	Specification/Allocation
<b>Hardware</b>		
Computing Resources	CPU/GPU specifications, number of cores	e.g. T4 GPU or equivalent for model training and inference
Memory	RAM specifications	e.g., 8 GB RAM for processing and training large datasets
Storage	Disk space for data, models, and logs	e.g., 1 TB SSD for storing datasets, trained models, and logs
<b>Software</b>		
Frameworks	Python frameworks	e.g., Flask for deployment
Libraries	Additional libraries	e.g., scikit-learn, pandas, numpy, matplotlib, seaborn for model training and evaluation

Development Environment	IDE, version control	e.g., Jupyter Notebook, Git
<b>Data</b>		
Data	Source, size, format	e.g., Kaggle dataset UCI Dataset