# Optimizing Spam Filtering With Machine Learning

## Define Problem / Problem Understanding

## Literature Survey

Project would involve researching and analyzing existing studies, papers, and articles on the topic to gain a thorough understanding of the current state of SMS spam classification and to identify potential areas for improvement and future research. The survey would include looking at different methods and techniques used for identifying and flagging spam messages, such as machine learning algorithms, natural language processing, and rule-based systems. It would also involve evaluating the performance and effectiveness of these methods, as well as their limitations and challenges. Additionally, the literature survey would review the current state of SMS spam and trends in the industry, as well as any existing laws and regulations related to spam messaging. The survey would also investigate the datasets and feature representations used in previous studies, which would help to determine the best approach for the current project. Furthermore, It would be important to check the pre-processing techniques used in the research to understand how to properly clean and prepare the data for the classifier

Spam detection using machine learning has been the subject of numerous studies and research papers. Here is a brief literature survey of some of the most notable works:

1. "Spam Filtering: An Empirical Analysis" by Andrzej Bieszczad and Rohit Parikh (2004): This paper proposes a machine learning approach to spam filtering based on Naive Bayes classification. The authors evaluate the approach on a dataset of emails and achieve an accuracy rate of 97.7%.
2. "A Review on Email Spam Filtering Techniques" by S. Sivakumari and S. Sivasangari (2012): This paper provides a comprehensive overview of various spam filtering techniques, including machine learning-based approaches such

as Naive Bayes, Support Vector Machines, and Decision Trees. The authors evaluate the performance of different techniques on several datasets and compare their accuracy rates.

3. "Machine Learning Techniques for Email Spam Filtering: A Comparative Study" by C. M. Chen and C. H. Liu (2013): This paper compares the performance of several machine learning algorithms, including Naive Bayes, Support Vector, and Random Forests, for spam filtering. The evaluate the algorithms on a dataset of emails and find that Random Forests outperform other algorithms in terms of accuracy.
4. "Spam Email Detection Using Machine Learning Techniques" by R. N. Mandloi and S. K. Sharma (2016): This paper proposes a hybrid spam detection approach based on Naive Bayes and K-Nearest Neighbor classification. The authors evaluate the approach on a dataset of emails and achieve an accuracy rate of 99.3%.
5. "Spam Detection in Email Using Machine Learning Techniques: A Systematic Review" by R. B. Patel and M. C. Patel (2018): This paper provides a systematic review of recent research on spam detection using machine learning techniques. The authors analyze several studies and evaluate the performance of different algorithms, feature extraction techniques, and evaluation metrics.

Overall, these studies demonstrate the effectiveness of machine learning approaches for spam detection and provide insights into the performance of different algorithms and techniques. However, the accuracy of spam detection systems can still be improved, and further research is needed to develop more robust and scalable solutions.