Optimizing Spam Filtering With Machine Learning

Define Problem / Problem Understanding

Specify the Business Problem

Over recent years, as the popularity of mobile phone devices has increased, Short Message Service (SMS) has grown into a multi-billion dollar industry. At the same time, reduction in the cost of messaging services has resulted in growth in unsolicited commercial advertisements (spams) being sent to mobile phones. Due to Spam SMS, Mobile service providers suffer from some sort of financial problems as well as it reduces calling time for users. Unfortunately, if the user accesses such Spam SMS they may face the problem of virus or malware. When SMS arrives at mobile it will disturb mobile user privacy and concentration. It may lead to frustration for the user. So Spam SMS is one of the major issues in the wireless communication world and it grows day by day.

The spammers have focused their attention on sending spam through short messages services (SMS) to mobile users. The user has confidential and personal information such as passwords, images, numbers of credit card, contact lists that stored on these phones, making those users more vulnerable to cyberattacks by spam SMS. Spam may leak sensitive information, privacy invasion, or access unauthorized information. Spammer are people with unethical activities can access data in smartphone without the end-user knowledge, exposing the privacy of the user to the path that results in financial or functional cost. They have had some success because of the lack of appropriate tools to deal with spam messages. This project is used to review and study the relative strengths of various emerging technologies to detect spam messages sent to mobile devices. Machine Learning methods and modeling techniques have been remarkably effective in classifying spam SMS. Detecting SMS spam suffers from a lack of the availability of SMS

dataset and a few numbers of features in SMS. Various features extracted and dataset used.