**Methods for Detecting Cyber Attacks**

**SMS phishing detection based on “Normal” traffic learning**

**Ariel University**

**Introduction**

SMS Spam messages are any type of unwanted or harmful messages, such as advertisements, frauds, business services, etc...

Message filtering can range from a simple static list of prohibited terms, to advanced machine learning systems that constantly adapt based on the messages passing through them.

Spam detetction is a Supervised Machine Learning problem which normally uses the following methods of classifying wether a message is a ham or spam:

Naïve Bayes: it is considered one of the

* Naïve Bayes: considered the simplest classification method.
* Random Forest
* Logistic Regression
* Support Vector Machine (SVM)

SMS filtration services fits well with supervised machine learning algorithms as a classification tool. Supervised ML works by being fed by trained data labels and then it gains the ability to predicts labels on new data.

**Work Process:**

1. Through using Kaggle Dataset (SMS Spam Collection Dataset) which contains tagged messages with 0 for spam and 1 for ham, we will build and train our own model classification model. The model’s goal will be to determine which messages are malicious.
2. The dataset will be split into training data (70% at the beginning) and testing data (remaining 30%)
3. We will then compare the model’s performance against different types of classification algorithms.
4. Running our model vs the <https://github.com/bit-ml/date> model.
5. Additional consideration might be taken if the model performs poorly, such as adding more data, changing the training\testing data ratio, adding more steps to the calculation of the error function, etc…