# E-mail Classification System for Bank Internal mail System

K.T Atapattu

139155D



Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Master Degree of Science in Information Technology.

**April 2016**

# Declaration

I hereby certify that this project and the all the artefacts associated with it is my own work and it has not been submitted before nor is currently being submitted for any other degree programme.

**Full name of the student:** Kasun Tharanga Atapattu

**Student Number** **:** 139155D

**Signature of the student**: ……..... Date: ……



**Name of the supervisor:** Mr. Saminda Premaratne

**Signature of the supervisor**: ……..... Date: ……

# Dedication

This Dissertation is dedicated to my loving parents for being part of me and encouraging me always being by my side.

# Acknowledgement

Apart from the efforts of me, the success of any project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this Research.

Special mention goes to my supervisor, Mr. Saminda Premarathne not only for his tremendous academic support, but also for giving me opportunity to successfully complete the research.

My next thank goes to Prof. Asoka S Karunanda who taught us Research Methodology and Literature Review and Thesis Writing subjects which were more benefited us throughout the research work

It is my great pleasure to thank all the other Senior lecturers, Lecturers, Instructors, and staff members who helped us in many ways to make this Research. The guidance and support received from all the members who contributed and who are contributing to this project, was vital for the success of the Research. I am grateful for their constant support and help.



Then I would like to thank my all the batch mate of the M.Sc Information Technology batch 7 in faculty of Information Technology for their various help and support. All so other friend of the faculty as well as friends for outside who gave me supporting and encouraging us with their best wishes.

# Abstract

My project is an implementation of 'Naive Based Algorithm' to classify E-mails. The main idea of the project is to implement algorithm in a effective manner to classify bank E-mails in bank internal authenticated mail system. E- mail has the potential to improve efficiency and reduce costs involved in communication. Even after the advent of newer technologies such as instant messaging and VoIP, email remains the number one application for business communication.

With the increasing of information on the internet based communication, our bank internal authentication mail system needs an efficient tool to classify the E-mails into categories. In this way, we can easily classify E-mail from large amount of E-mails available. Automated text categorization is a process that assigning pre-defined category labels to E-mail based on the contents.

Text categorization has many applications. For example, we can classify web pages into different categories to speed up the internet search, which is very useful for some search engines like Yahoo, Google etc. Also E-mail service providers are using those classification techniques for spam filtering and E-mail classification as well.

I have trained the develop algorithm from more than thousand pre - categorized E-mails. I have tested the text categorization algorithm developed based on naive based classifier with several different size data sets. Accuracy is evaluated as well. Experiment results shows my conclusion is efficient.

# Contents