

CATHOLIC UNIVERSITY OF BUKAVU



B.P.285 BUKAVU

FACULTY OF SCIENCES

Department of Computer Science

Development of a messaging application for communication and detection of spam on a mobile network, case study of Airtel, Vodacom and Orange.

Presented by : **MURHULA BYABUSHI Christian**
*Dissertation presented and defended in order to obtain the
degree of Bachelor in Computer Science.*

Option: Network and Telecommunications
Degree: Final year of Bachelor

Supervised by: ***Hw.* MUGISHO MUSHEGERHA Youen**
Directed by : ***Ph.D.* Elie ZIHINDULA**

Academic year: 2022-2023

Contents

Introduction	4
0.1 Context and generalities	4
0.2 Problematic	4
0.3 Hypotheses	5
0.4 Delimitation and objectives	5
0.4.1 Delimitation	5
0.4.2 Objectives	5
0.5 Interest	5
0.6 Research Methodology	6

List of Figures

List of Tables

Introduction

0.1 Context and generalities

With the increasing of use of mobile devices and the rise of mobile communication, the number of text messages sent every day has grown exponentially. According to *Statista*, a company that provides market and consumer data on a wide range of topics, including digital media and technology; the number of mobile messages sent worldwide in 2020 was 3.5 trillion [8]. Along with this growth, there has been an increase in the number of spam messages that aim to deceive people into providing personal information, sending money, menacing to death or taking other actions that benefit the scammer.

To address this problem, the development of a messaging application with advanced spam detection capabilities is essential. This dissertation focuses on the development of such an application for mobile networks, with a specific case study of Airtel, Vodacom and Orange.

Indeed, Airtel, Vodacom, and Orange are major telecommunications operators which mostly function in Democratic Republic of Congo, with a significant market share and importance in online services, including messaging.

0.2 Problematic

In communication domain, we use mobile devices or phones for sharing *SMS*, Email, chats by using some specific apps. Among all these we use specifically *SMS* for personal and professional information sharing [6]. The SMS stands for Short Message Service, which is a text messaging service for mobile phones and other mobile devices. It allows users to send and receive short messages of up to 160 characters [7]. It also possible to send or receive automatic SMS which are not sent by human, whereas by using web interface or API [5].

Due to all these means, we are facing against malicious attacks inside of messages which are commonly responsible of annoyance of users, wasting mobile network resources and potentially introducing security risks like *Simjacker* vulnerability which was discovered in 2019, which allows attackers to exploit a weakness in the SMS messaging system to remotely install spyware on mobile devices [3]. Besides, the spam messages are increasing likelihood of users by leading them to ignore important messages or being more hesitant to engage with *SMS* marketing campaigns [2].

Considering all above issues caused by spam, what are key technical challenges that could be addressed in messaging apps which can effectively facilitate the communication as well as detecting and filtering spam messages ?

0.3 Hypotheses

Using the content-based filtering techniques, which involves analyzing the content of messages and determining it is spam or not. This would be done by utilizing the Machine learning algorithms which are : **Naive Bayes, Logistic Regression, and Support vector Machines**. All these would be combined by the ensemble methods for making a more predictive model [9].

Besides, we would integrate the model inside of the system able of adding technically in a blacklist or whitelist suspect users based on the specific probability of being a potential attack.

0.4 Delimitation and objectives

0.4.1 Delimitation

The present work aims to develop a messaging application for communication and detection of spam on a mobile network, with a specific focus on the case of Airtel, Vodacom or Orange in Democratic Republic Of Congo(DRC). The study will be conducted over a period of nine months, from January to November 2023.

0.4.2 Objectives

This system present 2 types of objectives such are: functional and non functional.

As functional objectives, this system consist of developing a messaging application that can facilitate efficient communication between users; implementing the machine learning models which has the capacity of classifying the messages and preventing spam messages under a certain probability; designing and integrating a user-friendly interface for messaging application;

For non-functional objectives, it allows the user that the messaging application complies with relevant privacy laws and regulation to protect user data information, reducing the attacks and frauds; Increasing the trustfulness of users and mobile services compaignie provider, optimizing the power resources of the user.

0.5 Interest

Personally, this thesis allows the author to gain knowledge and more experience in the field of mobile networks and messaging applications.

Socially, the developed system would facilitate communication and reducing the impact of spam messages, which can cause annoyance and stress to citizens.

Economically, this system of detection helps to save business server time and resources by filtering out spam messages allowing for more targeted marketing efforts.

Scientifically, the research achieved contribute to the advancement of the science in the field of Mobile Networks, SMS messaging, and Machine Learning. Especially the combination of the 3 machine learning 's models implemented is a contribution considerable for researchers.

0.6 Research Methodology

Throughout this thesis, the research methodology will be used to guide the study towards achieving its objectives. The research will adopt a descriptive research design to describe the development of a messaging application for communication and detection of spam on a mobile network. The study will focus on both qualitative and quantitative research methods [4]. The qualitative method will involve a literature review, interviews and analysis of collected messages. while quantitative method will focus on the development and testing of the messaging application.

The research will be conducted in two phases. The first phase will involve data collection through a survey questionnaire that can be completed on website, or can be directly provided to the web interface (API) by the mobile phone users for collecting their experience with messages and especially spams. Thus, the data collected will be analyzed using descriptive statistics [1] to identify the common types of spam messages and the frequency of occurrence, languages inside, and other attributes.

The second phase will involve the development of the messaging application using the data collected from the survey and the analysis of existing messaging applications. The development of the application will be guided by the principles of agile software development by using Python and Django framework for *Back-end* and HTML,CSS and JavaScript for *front-end*. Then,the application will be tested using various machine learning models which include Naive Bayes, Logistic regression, and Support Vector Machine.

The evaluation of the messaging application will be conducted using both quantitative and qualitative methods. The qualitative evaluation will involve the measurement of the application's accuracy and efficiency in detecting and filtering spam messages, will the qualitative evaluation will involve a user study to determine the usability and user experience of the application.

Bibliography

- [1] Allan Bluman. *Elementary statistics: A step by step approach*. McGraw-Hill Education, 2017.
- [2] Guangquan Chen, Weijun Wang, and Xuan Zhou. A survey on sms spam filtering techniques. *Journal of Network and Computer Applications*, 80:149–159, 2017.
- [3] Catalin Cimpanu. Simjacker vulnerability exploited for surveillance by at least one nation-state. *ZDNet*, 2019.
- [4] John W Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2014.
- [5] Sunil Kumar Jangir, Manoj Kumar Sharma, and Pawan Kumar Gupta. Design and implementation of sms gateway api for mobile communication networks. *International Journal of Computer Applications*, 151(9):1–5, 2016.
- [6] M Lavanya and KR Aruna. Sms spam detection using deep learning. *Journal homepage: www.ijrpr.com ISSN*, 2582:7421.
- [7] Gwenael Le Bodic. *Mobile messaging technologies and services: SMS, EMS and MMS*. John Wiley & Sons, 2005.
- [8] Statista’s own team of researchers and analysts. *Number of mobile messages worldwide from 2019 to 2023 (in trillions)*. <https://fr.statista.com/>, 2020.
- [9] Sebastian Raschka and Vahid Mirjalili. Python machine learning: Machine learning and deep learning with python. *Scikit-Learn, and TensorFlow. Second edition ed*, 3, 2017.