



FINAL YEAR PROJECT REPORT

**ANOMALY DETECTION IN CELLULAR
NETWORKS USING MACHINE LEARNING**

**In fulfillment of the requirement
For degree of
BS (COMPUTER SCIENCES)**

By

SYED TALAL HASSAN

51800 BSCS

MUHAMMAD MURTAZA KHAN

51831 BSCS

BILAL DOST MALIK

51812 BSCS

SUPERVISED

BY

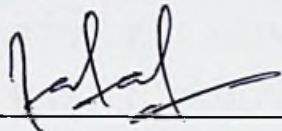
SIR BILAL MUHAMMAD IQBAL

BAHRIA UNIVERSITY (KARACHI CAMPUS)

SPRING-2021

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Signature : 

Name : SYED TALAL HASSAN

Reg No. : 51800

Signature : 

Name : MUHAMMAD MURTAZA KHAN

Reg No. : 51831

Signature : 

Name : BILAL DOST MALIK

Reg No. : 51812

Date : 17/06/2021

The copyright of this report belongs to Bahria University according to the Intellectual Property Policy of Bahria University BUORIC-P15 amended on April 2019. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this report.

© 2019 Bahria University. All right reserved.

ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express my gratitude to my research supervisor, Sir Bilal Muhammad Iqbal for his/her invaluable advice, guidance and his/her enormous patience throughout the development of the research.

In addition, We would also like to express my gratitude to our loving parents and friends who had helped and given me encouragement.

ANOMALY DETECTION IN CELLULAR NETWORKS

ABSTRACT

The identification of "abnormal" events in datasets has long been a focus of machine learning research. Anomaly detection or outlier detection are terms used to describe this technique. Grubbs provided the most likely first definition in 1969. Outliers are observations that appear to diverge significantly from the rest of the sample. Although this criteria remains relevant today, the reason for spotting outliers has changed dramatically. Because pattern recognition algorithms were quite sensitive to outliers in the data back then, the major reason for the detection was to eliminate the outliers from the training data thereafter. Data cleaning is another name for this operation. The interest in anomaly detection has waned significantly with the emergence of more robust classifiers. However, in the year 2000, academics began to become increasingly interested in the anomalies themselves, as they are frequently linked to certain noteworthy occurrences or questionable data sets. Anomaly detection methods are currently employed in a wide range of applications and are frequently utilised to supplement classic rule-based detection systems.

Traditionally, the design of a cellular network has focused on energy and resource optimization to ensure that the network operates smoothly even during peak hours. This means, however, that radio resources are frequently overprovisioned in cells. In order to react to changing user demands in the most effective way possible in terms of energy savings and frequency resource use, next-generation cellular networks require dynamic management and configuration. Machine Learning approaches are now being studied in mobile networks to assist with resource management. In this scenario, you will look at how machine learning can be used to detect abnormal network usage patterns that might lead to a change in the base station's setup.

TABLE OF CONTENTS

INTRODUCTION	12
1.1 Background	12
1.2 Problem Statements	12
1.3 Aims and Objectives	13
1.4 Scope of Project.....	13
2 LITERATURE REVIEW.....	15
2.1 Anomaly Detection Methods.....	15
2.2 Rule Based Method	15
2.3 Model Based Approach	16
2.4 Static Single Processing	17
2.5 Clustering Based Methods.....	17
2.6 Information Theory Methods	18
2.7 Root Cause Analysis	19
3 DESIGN AND METHODOLOGY	20
3.1 Design.....	20
3.2 Data Gathering	20
3.3 Data Pre Processing.....	20
3.4 Problem with high dimensional data	21
3.5 The Curse of Dimensionality	21
3.6 Visualising the high dimensional data.....	21
3.6.1 Principal Component Analysis.....	22
3.6.2 t-SNE.....	23
3.7 Feature Filtering	23
3.8 Approaches for Anomaly Detection.....	23
3.8.1 Clustering based method for anomaly detection.....	23

3.8.2	Decision Trees.....	24
3.9	Training the classifier.....	24
3.10	Feature Extraction.....	25
3.11	Feature Importance.....	25
3.12	Evaluation Matrix.....	26
3.12.1	Confusion Matrix.....	26
3.13	Methodology.....	27
3.14	Information Gathering.....	27
3.15	Designing.....	27
3.16	Development.....	27
3.17	Testing.....	27
3.18	GUI.....	28
3.18.1	PCA.....	28
3.18.2	t-SNE.....	29
4	IMPLEMENTATION.....	30
4.1	Dataset.....	30
4.2	Components in project.....	31
4.3	Data Pre-Processing.....	32
4.4	Data Visualization.....	32
4.5	Training a Model.....	33
4.6	Feature Reduction.....	34
4.7	Setting Threshold.....	34
4.8	Feature Importance.....	35
4.9	Evolution Of Error For Different Threshold.....	36
4.10	Confusion Matrix.....	37
4.11	Best Performance Feature Reduction.....	37
5	RESULTS AND DISCUSSIONS.....	39

5.1	Theory	39
5.2	Neural Network	39
5.3	Evaluation Metrics	40
5.3.1	Confusion Metrics	40
5.4	Result.....	40
6	CONCLUSION AND RECOMMENDATIONS.....	42
6.1	Conclusion.....	42
6.2	Research Analysis	42
6.3	Process.....	43
6.3.1	Data Processing.....	43
	REFERENCES.....	44