SOCIOECONOMIC MAPPING USING MOBILE CALL DETAIL RECORDS FOR SRI LANKA

Chandima Dileepa Rajaguru

(168258P)

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

February 2020

SOCIOECONOMIC MAPPING USING MOBILE CALL DETAIL RECORDS FOR SRI LANKA

Rajaguru Mudiyanselage Chandima Dileepa Rajaguru

(168258P)

Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science and Engineering

Department of Computer Science and Engineering

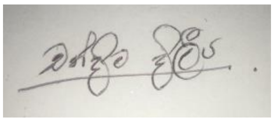
University of Moratuwa
Sri Lanka

February 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works.



2020-05-30

....

Chandima Dileepa Rajaguru Date

The above candidate has carried out research for the Masters dissertation under my supervision.

....

Dr. Amal Shehan Perera Date

Abstract

CDR (Call Detail Record) is a data record that is generated by a telephone exchange or telecommunication equipment which contains details of that telephone call. These records are utilized by telecommunication service providers for their billing purposes. High volume of data generates in quick time which contains customer specific data with temporal and geographic information. Other than CDR data, telco systems have various data sources such as customer payment data and device information. Telco service providers collect CDR and store them for a limited period of time for various activities. It can be repurposed other than billing activities.

CDR data can denote various aspects of human behavior such as human relationships, expenditure power and mobility. Those aspects can help governance of the country regarding economic development and resource allocation in timely manner. In this research, CDR data records were integrated with other telco data sources in order to analyze and predict the economic behavior of a specific geographical area in Sri Lanka.

Big data and Machine Learning techniques were used to extract the customer behavior from CDR data. Big data processing techniques were applied on CDR data and telco data sources in order to identify properties of customers in a specific geographic area over a time period. Then those identified properties were evaluated to see whether they reflect the economic behavior in that area or not. After identifying dominant features related to the economy, Machine Learning techniques were applied on them to see the feasibility of predicting the economic behavior in the targeted area. The results were evaluated and interpreted as a part of this research. Such results will be very useful for the governance in order to understand the economic conditions in a specific geographical area and make the policies to address poverty over the time.

ACKNOWLEDGEMENT

I would like to offer my sincere gratitude to my family for the continuous support and motivation given to me to make this dissertation a success. I also express my heartfelt gratitude to Dr. Amal Shehan Perera, my supervisor, for the supervision and advice given throughout this research for finishing it successfully. Then I pay my gratitude Mobitel (Pvt) Ltd for the guidance and facilities given to me in order to carry out this research. I also thank my friends who supported me in this whole effort.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CDR	Call Detail Record
ML	Machine Learning
DSD	Division Secretariat Division
GND	Grama Niladhari Division
OLAP	Online Analytical Processing
RDBMS	Relational Database Manage System
SEL	Socio Economic Level
MPI	Multidimensional Poverty Index
DHS	Demographic and Health Survey
BTS	Base Transceiver Station
PCA	Principal Component Analysis
SVM	Support Vector Machine
DCS	Department Census and Statistics