IMPROVING UNLOADING TIME PREDICTION THROUGH DRIVER AND CUSTOMER SEGMENTATION

Liyadipita Appuhami Mudiyanselage Ranula Prasanna Bandara Liyadipita

(209352E)

Master of Science in Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

July 2022

IMPROVING UNLOADING TIME PREDICTION THROUGH DRIVER AND CUSTOMER SEGMENTATION

Liyadipita Appuhami Mudiyanselage Ranula Prasanna Bandara Liyadipita

(209352E)

Thesis/Dissertation submitted in partial fulfilment of the requirements for the degree Master of Science in Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

July 2022

DECLARATION

I declare that this is my work and this dissertation does not incorporate without acknowledgment any material previously submitted for the 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 acknowledgment is made in the text.

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

Signature: Date:

The above candidate has carried out research for the Master's dissertation under my supervision.

Signature of the supervisor: Date:

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to all those who provided support to make my research on "IMPROVING UNLOADING TIME PREDICTION THROUGH DRIVER AND CUSTOMER SEGMENTATION" successful.

First of all, I would like to express my gratitude to my project supervisor Dr. Uthayashanker Thayasivam, Senior Lecturer, Department of Computer Science and Engineering. I am highly indebted to him for his guidance and constant supervision as well as for providing necessary information regarding the project and for his support in completing the project successfully.

I am sincerely thankful to the final year project coordinator Dr. Charith Chittaranjan, Senior Lecturer, Department of Computer Science and Engineering for the support given throughout the project time period. Further, I would like to extend my gratitude to Dr. Nisansa De Silva for participating in evaluations and providing me with very useful guidance to make my research successful.

Specially I would like to thank Prof. Indika Perera, Head of Department, Department of Computer Science and Engineering for his assistance and coordination to conduct the research without any issues during the final year.

Finally, I wish to thank the academic and non-academic staff of the Department of Computer Science and Engineering and colleagues for the support and encouragement given.

ABSTRACT

Modern-day society is driven by transportation networks. Now it is easier than ever to order your daily necessities through online platforms. The research interest in this thesis focus on the delivery aspect lies with ordering. A successfully completed delivery means a properly addressed vehicle routing problem. The data set that is involved in the study refers to a large amount of perishable good cases that are delivered through large trucks. Each truck caters to 8-10 customers in a day. Since these are perishable goods delivered to people in the foodservice industry, they expect a sound ETA of their delivery to plan ahead for meal preparations.

To provide an ETA in a multi-stop route there are two variables to be solved. One is the travel time between stops, which modern-day map services would output without a hassle. However, the next important thing is the unloading time needs to calculate with the historical data. The study suggests a way to involve customer profiling and driver profiling so that unloading time prediction can be done with those two variables along with the delivery volume of the stop.

Modeling these two variables into a regression model was a challenge on its own due to their large dimension of them. Segmentation of the said variables and using segment mean yielded better results in regression compared to using a label encoding technique blindly which introduced an orderly nature to features from the id itself. Furthermore, once segment means were clustered based on their distribution and provided a cluster identifier that justifies the orderly nature, models were able to yield their least MSE.

Finally, this study highlights the importance involving of the customer site and the driver's experience in the unloading time. Also, this study has presented a way of representing such variables with a high cardinality in a meaningful manner so that model can be built with less error. This will provide a good starting point for further analysis on similar research interests in the future

TABLE OF CONTENTS

DECLARATION i
ACKNOWLEDGEMENTS ii
ABSTRACT iii
LIST OF FIGURES vi
LIST OF TABLES vi
1 INTRODUCTION 1
1.1 Research Problem 1
1.2 Challenges 2
1.3 Research Objectives 2
1.4 Contributions of Research 3
2 LITERATURE REVIEW 4
2.1 Vehicle routing problem 4
2.1.1 The problem and the solving approaches 4
2.1.2 Parameters involved in VRP 5
2.2 Unloading time prediction 8
2.3 Representing Categorical Variables 10
2.4 Regression models and evaluation techniques 13
2.5 Avoid overfitting in regression models 13
3 METHODOLOGY 15
3.1 Creating a dataset 15
3.2 Dataset Description 16
3.3 Data Pre-processing and Preparation and Feature Selection 18
3.3.1 Analysing the dataset and preprocessing 19
3.3.2 Feature Selection 21
3.4 Categorizing customers and drivers into segments 21
3.4.1 Label Encoding with ordinal coding 22
3.4.2 Using mean unloading time of a segment 22
3.4.3 Segmentation of means and use of ordinal label encoding 23
3.5 Training ML models 25
4 EVALUATION 26

4.1 RMSE 26
4.2 R square 26
4.3 Accuracy 26
5 RESULTS 28
5.1 featureSet1 28
5.2 featureSet2 28
5.3 featureSet3 29
6 DISCUSSION 30
7 CONCLUSION 31
8 REFERENCES 32

LIST OF FIGURES

Figure 1: Unloading time calculation equation 9
Figure 2: Unloading time calculation coefficient equation 9
Figure 3: Methodology Steps 15
Figure 4: CSV generation steps 16
Figure 5: GPS unloading time VS delivery cases 19
Figure 6: Algorithmic unloading time vs delivery cases 19
Figure 7: GPS unloading time VS delivery cases - Without Outliers 20
Figure 8: Algorithmic unloading time vs delivery cases – Without Outliers 20
Figure 9: Pearson correlation Plot 21

LIST OF TABLES

Table 1:VRP constrains 5
Table 2: VRP Parameters 7
Table 3: Encoding Techniques 11
Table 4: Dataset Description 17
Table 5: Label Summary 22
Table 6: Driver and Customer Overall Segment Details 23
Table 7: Distribution Information of Mean Unloading Time of a Delivery Case For Customer 24
Table 8: Distribution Information of Mean Unloading Time of a Delivery Case For Driver 24
Table 9: Results for featureSet1 28
Table 10: Results for featureSet2 28
Table 11: Results for featureSet3 29