Chapter 1

Introduction

The following section will give an overview of the research topic chosen. This section will

include Research Background, Research Purpose, Hypothesis, and Research Questions,

Significance of Research, Importance of Research, Research Boundaries and Research

Outline.

1.1 Research Background

The main area of study proposed is Machine Learning. Machine learning is a branch of artificial intelligence which is centered on the concept that system can learn from data, gain the ability to identify patterns, thus leading to decisions with minimal human intervention. The iterative position in machine learning is crucial as models are subjected to new data, they adapt in an independent matter (Machine Learning: What it is and why it matters, 2021).

In this research, the research will focus on supervised machine learning. Super- vised machine learning is the pursuit for algorithms that use externally supplied occurrences to reason and create a general hypothesis, which then make expectations for future instances. The goal in supervised learning is to construct a short model of the distribution of group labels in predictor feature conditions. The classifier of result is then used to designate group labels in the testing occurrences where the values of the predictor features are recognized, yet the significance of the group level is unknown (Kotsiantis, 2007).

The researcher intends to focus his research on predicting auto dealer car sales.

The research method for this research will be quantitative, as the data will collected from several auto dealers through an online survey and experiment. The experiment will make use of a past dataset from the internet as this will not be given from the dealers themselves. The data collected will then be analyzed via descriptive and inferential statistics techniques.

1.2 Research Purpose

The motivations of this research are trifold which contributes to the researcher’s personal motivation for the choice of the research theme: Firstly, the researcher has a great interest in the automotive industry alongside the understanding of market fluctuation. As a car enthusiast, the flexibility in prices of newly developed cars is of interest as it may possibly:

• Be a possibility of a future purchase.

• Benefit the organization and lead to more research in improving automobiles, improved and more sustainable production, and ultimately create an automobile of personal interest in long term or

• Detriment the profitability of the organization leading to less production, more costs incurred, and ultimately failing to create personal interest in their automobiles.

Secondly, machine learning is a technique of data evaluation which presets analytical model constructing. The concept is centered around the system can learn from data, obtain the ability to identify patterns, thus judgments are made with minimal human intervention.

Finally, Machine Learning has made it possible to produce models capable rapidly and automatically of larger and more complex data analysis, faster delivery, and more accurate results. With the construction of accurate models, businesses have an increased chance of identifying profitable chances and avoiding unknown

risks (Machine Learning: What it is and why it matters, 2021). The application of machine learning has enhanced many individual studies such as neural networks, genetic learning, cancer prediction and prognosis, drug discovery and development, ecological modelling, animal behavior studies, and so on.

1.2.1 Hypothesis and Research Questions

The Hypothesis of this research is: Supervised machine learning, can predict auto dealer car sales.

The research questions in this study are:

RQ1: Does applying supervised machine learning technique increases auto dealer car sales?

RQ2: Can machine learning using a supervised learning method accurately predict auto dealer sales?

This section will present a detailed idea about how the proposed study will be conducted. This section will include: Target Participants and Sample size, Data Collection Methods and Data Analysis.

The target participants are vehicle auto dealers. In this case with a population size of 116 auto dealers (Car Dealers in Malta Gozo, 2021), a 10% margin of error, and a 95% confidence, the sample size is of 53 respondents (Sample Size Calculator by Raosoft, Inc., 2021). This target group has been selected as it fit due to its relevance to the study in confirming the information discovered.

Online surveys to gather information from Auto dealers. Topics for the survey would be regarding the prices, market research, if such a system would be ideal for their use, would they suggest it etc. This will take place online. The experiment for

this study will consist of understanding the idea on why decisions were made, the demographics behind these decisions, and the outcome. Upon receiving instances of these, the system will learn and adjust to scenarios, gaining capability of prediction in auto dealer sales.

The data analysis will consist of Chi squared, Cross-tabulation, T-test, and Anova. Chi squared, which is commonly used to assess tests of independence when using Cross tabulation. The test of independence assesses whether there is an association between the two variables via witnessing a repetition of reactions expected if the variables were truly objective of one other (Using Chi-Square Statistic in Research - Statistics Solutions, 2021). The T-test being an inferential statistic, it shall be used to determine any substantial difference between two group means, potentially related in certain features (T-Test Definition, 2021). Anova splits observed cumulative variability observed inside a data set into two parts: systematic factors and random factors. Systematic factors have a statistical influence on the data set given, while random factors do not (How Analysis of Variance (ANOVA) Works, 2021).

1.2.2 Significance of this Research

The research will be carried out to create a machine learning environment in which car dealership prices and sales in Malta are analyzed and learned. Ultimately this would lead to the machine predicting the future sales of car dealership cars, with the use of supervised regression learning. This will aid risk management and provide advantages and disadvantages of important business decisions, as well as understanding the organization’s current situation and future situations to make the correct business decisions.

Dealerships are prone to interest themselves in such a product due to the advantages in aiding future business decisions. Finally, this study will also compare the advantages and disadvantages of making use of such a system, while also comparing risks and whether it is worth taking them.

Furthermore, this research can be used as a reference by other researchers’ conducting similar studies to obtain the applicable and justifiable resource about the use of supervised machine learning or it can be also used as a reading material for any person who is interested in this area of study.

1.3 Importance of Research

The contributions made by this study will involve additional insight to the effects of implementations and understanding of future new car markets from auto dealers in Malta. This study will help the understanding of the upcoming new car sale sales and the decision changes auto dealers are prone to with the use of machine learning prediction. Prices are constantly fluctuating, and, in some instances, second-hand cars are more expensive than new cars due to an inflation which has been caused by popularity, rarity, manufacturers concluding production, etc. The Artificial intelligence will learn through a series of examples the patterns while taking into consideration the effects of both independent and dependent variables, estimate an accurate price point which will then reveal which factors are imperative, what changes or precautions can be made, and what data can be altered. Further insight into this topic will avoid future problems, understand if machine learning predictions are worth the effort and price and if this will be profitable.

1.4 Research Boundaries

This study will remain within the supervised machine learning environment. A supervised learning model has been chosen for omitting manual classification work and for making predictions based on labeled data in the future. Ideally the use of supervised machine learning is in cases where both input and the corresponding output are known, where labeled data is used in training the machine. The supervised learning method takes direct feedback to help understand whether the

correct answer has been predicted. Following the data input this machine learning method is to be supervised in order to generate a more accurate prediction.

The target market has been chosen due to the researcher’s interest in the auto- motive industry along-with the recognition of market fluctuation. This will aid in interpreting the flexibility in prices of newly developed cars, possibly to be a benefit of an organization, improving research on automobiles, and feasible production.

1.5 Research Outline

This research will contain the following chapters:

1. Introduction:

2. Literature Review:

I Types of Machine Learning

II Supervised Machine Learning

III Development Methods of Supervised Machine Learning

IV Benefits of Supervised Machine Learning

V Predictive Analysis Approach and Current Applications

VI Suggested Development Solution

VII Conclusion

3. Research Methodology:

4. Analysis of Results and Discussion:

5. Conclusions and Recommendations:

1.6 Conclusion

This section will give an overview of the literature that is related to the research topic

chosen. This section will include Machine Learning, Types of Machine Learning, Supervised

Machine Learning, Development Methods of Supervised Machine Learning, Benefits of

Supervised Machine Learning, Predictive Analysis Approach and Current Applications,

Suggested Development Solution.