University of Nevada - Reno
Department of Computer Science and Engineering
CS 425/625

“Exploring Data Mining in Used Car Data”

Team #8:
Patrick Austin, Eric Stutzman, Guang Xu, Mile Cilic

Instructors: Dr. Sergiu Dascalu and Dr. Devrin Lee

External Advisors: Dr. Lei Yang, UNR and Stephen Williams, Marketing Evolution

16 October 2017

Abstract

Data analysis on large datasets has been a major topic of research in recent years, and it has been shown to be beneficial for enterprises which need to advertise their products or services. Historically, the ability of businesses to target advertising to specific groups has been limited, but advances in data mining and machine learning technology have enabled businesses to maximize advertising effectiveness by speaking directly to interested customers. On the other hand, it is difficult to use data mining algorithms effectively on very large and multidimensional datasets, and it is also difficult to extract results into useful information in a human-readable format. We present a project enabling new marketing solutions for used car dealers aiming to increase their sales and reduce their advertising costs. Our solution is based on performing data mining classification algorithms on a 50GB used car transaction history dataset made available to us by the analytics company Marketing Evolution.

Project Description

Main Goals and Objectives

The goal of our project is to explore a large dataset of used car sales provided by the company Marketing Evolution in order to discover information that could be used to provide data analytics services to customers. Namely, Marketing Evolution specializes in identifying the target market and creating targeted advertising plans, which are essential elements of a successful marketing strategy. Therefore, we will be applying data mining techniques to achieve the extraction of useful patterns from the provided dataset.

Given the rarity of large untapped datasets and their potential usefulness to customers in the marketing and advertising areas, we believe this project could be of significant use in the area of marketing analytics.

Main Functionality and Characteristics

We plan to build a back-end website that will consist of algorithms to do data mining tasks such as classification. Specifically, users will be able to input characteristics about their customers, and in return, they will get targeted information about that customer, such as a set of cars that the customer is most likely to be interested in. Additionally, if time and resources allow, we hope to

implement a front-end application for both Android and iOS mobile platforms which users can use to easily query the back-end resources we plan to build.

Intended Audience

The project's intended audience will be businesses seeking analytics and marketing recommendations derived from this data, such as used car dealerships and other used-car sellers. It is also possible that this data could be adapted to buyers of used cars to assist them in the search for a new vehicle.

Key Usability Goals

The ability to provide targeted car recommendations can help businesses selling used cars to reach their customers with minimal advertising overhead. We intend to support the use of this data to create marketing plans that focus on smaller, targeted segments of the overall market. Our aim is to provide a back-end that can service queries, and be used to visualize data with minimal latency and maximal uptime and provide as many services. We intend to maximize the ability of users to visualize and interpret the relationships obtained by data mining in a human-readable way.

Potential for Further Development/Product Enhancements

The scope of this project can vary greatly depending on the difficulty encountered in the course of realizing the initial data mining task. After completing the project, there is the possibility to continue working with Marketing Evolution to enhance the project, further explore the data, and integrate our work into Marketing Evolution's software solution and business plans. We could also expand the project towards a more consumer and shopper oriented solution.

Challenges and Obstacles

One of the main challenges that our team will face is inexperience with the required technologies. Our team members have limited experience in the data mining field. We will need to learn new programming languages to achieve our goals, due to our intent to use Python and the lack of team experience with the language. Two of our team members are currently enrolled in the CS 491: Data Mining course at the UNR, so they will most likely take the lead on the data mining tasks. This will most likely require the other two members to take extra time to study data mining to be up to speed.

A technology problem for this project is the computing power needed to perform operations on such a large dataset. Marketing Evolution has indicated the full dataset could be as

large as 50GB. This would most likely preclude our team from running any operations on this large dataset with our personal computers. In our initial meeting with Marketing Evolution, they suggested using a small sample of the dataset to create an accurate model. Once we are confident in our prediction model, then we might be able to use more powerful servers via Amazon Web Services to work on the entire dataset.

Since the dataset has so many attributes, it might be difficult to find a specific customer for our product. Used car salesmen are a very small market, so we might need to focus on expanding the functionality of the website instead of the smartphone app. With a feature-rich website we could market our prediction capabilities to car dealerships, or aftermarket parts manufacturers that are looking for a smarter way to reach their target market.

Technology Description

Our initial plan is to use Scikit-learn to perform data mining tasks through Python. Python will also be used to make the website with the web framework Django. We also plan to use Xamarin will to develop a cross-platform mobile app using C#.

Team Overview

Patrick Austin is an undergraduate student in the Computer Science & Engineering program at UNR. He has a prior degree in Political Science, and aims to use his humanities training to assist the team in persuasively presenting and explaining the project. He is enthusiastic about studying data mining algorithms and how they should be tailored to suit this real-world dataset.

Mile Cilic is an undergraduate student at the University of Nevada, Reno majoring in Computer Science and Engineering, and minoring in Mathematics. He enjoys solving mathematical algorithms which will be beneficial for this project. Mile is enthusiastic about building a mobile application and wants to specialize in this area in the future.

Eric Stutzman is an undergraduate senior in Computer Science & Engineering at UNR. Eric has participated in undergraduate research in cybersecurity at the Florida International University, and the semantic web at IUT de Bayonne. He will use his ongoing training in data mining to ensure predictions are as accurate as possible.

Guang Xu is a graduate student of Computer Science & Engineering at UNR. He earned his bachelor of science degree from UNLV in the summer of 2017. Guang has been exposed to data mining theories and algorithms when he was an undergraduate student, and he also has

experience with mobile application design in iOS platform which could help in the front end design.

Advisory Overview

The main advisor for this project is Dr. Lei Yang who teaches data mining at UNR. Our main point of contact at Marketing Evolution is Stephen Williams who is a recent graduate of UNR.

Professional Growth

This project will expose our team to the real world of software development by working with an industry partner as well as a faculty advisor. We will be better prepared for employment after working with sensitive data that requires us to sign a Non Disclosure Agreement (NDA). Our resumes will be further improved due to our development of a system that uses real world data to produce a useful answer.

Market Potential

Market Analysis

The team is enthusiastic about the market potential and commercial promise for this project. This project is being developed in cooperation with and making use of data provided by Marketing Evolution, an award-winning data analytics company. [Marketing Evolution, web.] The team's software may be directly applicable to Marketing Evolution's core business of analyzing data, and targeting advertising to maximize return on marketing money spent.

Q2 2017 saw an all-time record in used auto sales, with average sale prices at an all-time high. [Edmunds, web.] Used auto sales are in high demand and the market is competitive. Therefore being able to target advertising to a finer degree, or to assist dealerships in recommending cars to shoppers, could help our business users to leverage these favorable trends to the highest degree possible.

Competitive Analysis

Recent years have seen a proliferation of used car sale resources for both consumers and sellers. These include numerous Blue Book websites offering used car sale data, and sites like Craigslist and eBay which connect used car buyers and sellers directly. At the same time,

businesses offering targeted advertising in a variety of areas have grown into tech titans; something like recommending used cars based on gathered data is very much in the wheelhouse of Google and Facebook.

Narrowly speaking, the potential for this project to be used by Marketing Evolution means that direct competitors are other businesses in the marketing, analytics, and return-on-investment areas. These could include prominent upstarts in the area such as Kapost, Lattice Engines, and Mixpanel, all of whom offer similar products to the software solution sold by Marketing Evolution. [Entrepreneur.com, web.]

More broadly, there is considerable indirect competition in the marketing area as applied to used cars. As our software and our findings could possibly be adapted to these areas as well, some consideration is due to the aforementioned Blue Book websites, car sales websites, and advertising giants like Google and Facebook. All of these and many more could claim some measure of expertise in leveraging data about used car sales for use in advertising, even if this is not their primary business area.

Competitive Advantage

The team is optimistic that their approach, which connects state-of-the-art data mining techniques learned in consultation with the team's faculty advisor with access to a massive and little-explored dataset provided by industry, can bring something novel to the table in the marketing and analytics area. This data is new, ripe for exploration, and not available to the general public.

Since the relationships existing in this dataset relating customers to their used cars are yet unknown, the team can bring a competitive advantage into play by using data mining techniques to discover causal relationships that competitors have not yet discovered. Marketing Evolution may be able to tout its unique insights when offering its services to clients in the used car business.

Time Worked

Our first group meeting was with Dr. Yang and Stephen Williams in a video conference that lasted 30 minutes. Our second group meeting was a working lunch where we brainstormed about initial topics for 2 hours. We had a third meeting lasting a half hour where we prepared a rough draft and our plans for the subsequent meeting with the teaching staff. We met with the teaching staff concerning our initial proposal for 20 minutes the next day. Patrick spent 1.5 hours writing the Market Potential section. Mile spent 1.5 hours writing the first half of the Project Description up to and including the Potential for Further Development/Product Enhancements.

Eric spent 2 hours writing the last half of the Project Description starting with Challenges and Obstacles, as well as this section. Guang spent 1 hour writing the Abstract. The team spent an additional hour proofreading and editing the proposal document for submission. In total, each team member spent at least 5 to 6 hours on topics related to the project proposal.

Works Cited

Forrester's Q4 2016 Wave Report (website). Retrieved from https://www.marketingevolution.com/wave

Four Marketing Analytics Tools That Are Shaping the Industry (website). Retrieved from https://www.entrepreneur.com/article/241534

Used Car Market Quarterly Report (website). Retrieved from https://www.edmunds.com/industry-center/data/used-car-market-quarterly-report.html