**Autix: Designing a Machine Learning Model to predict and maximize the value of modified and specialty vehicles**

**Marcos Munoz, Mehedi Toufiqe, Sayema Rahman, Yaseen Mohammed**

The University of Texas at Dallas | Fall 2023 Capstone Project

In this project, we will develop and validate a pricing model related to AUTIX’s curated value metric algorithm known as xEstimate. This algorithm will take in a car’s information such as maintenance history, performance modifications, cosmetic modifications, social sentiment, and some baseline information and it will give the user a price for their car based on the input received.

Normally when a car owner wants to sell their vehicle, most online websites and apps don’t consider some of the aforementioned variables, so the owner gets a price that for them does not reflect everything that they have done to their car. AUTIX wants to provide their users with a good and accurate algorithm that will provide a much better price for their vehicle that considers more aspects. We hope to achieve the project goal by using Python and some of its libraries to help make/improve a working model.

**TABLE OF CONTENTS**

Introduction 2

Resources 2

Key Roles 3

Communication Plan 3

Timetable 3

Evaluation 4

Conclusion 4

Contact Information 4

Resources & Sources 5

Appendix 5

# **INTRODUCTION**

The project we were tasked with for our capstone was to facilitate the creation of a marketplace for modded cars. The specific task we were given was to extract data on existing modded cars and their prices to create a machine-learning model to predict the price of new modded cars that would be added to the marketplace. Kelly BlueBooks is a similar platform to our project however it does not take into account the value of modding your car and customization. We plan to use LLMs to process the vast amounts of unstructured data into neat columns so it is easier for models to use that to accurately predict value for the cars.

# RESOURCES

We were given a spreadsheet named “profiles” with data about the vehicle, such as the owners, make, model, year, etc. from Andrew, who is the data scientist at the company. We will analyze and extract information from the data. We used the help of Dr. Norrit Chandra to figure out which method would work best for our project. We used the help of the Google search engine to research different methods that could work for analyzing the “history/awards/story” column of the data.

To help analyze the data, we used different machine learning models to help us extract the features from our data. We used a GPT API-based token model and the LLama2 LLM method. We used Google Colab to run our code with the provided data file. We all used different computers to run our programs on. The devices we are using are the LG Gram Pro, Lenovo Yoga, HP, and a MacBook Pro.

We were given a website named classiccars.com to use to extract and populate a spreadsheet with data about the cars listed on the website. We will be using this data as our training data for the “profiles” spreadsheet.

# **KEY** **ROLES**

For our group, we all decided that it would be best if we all tried to split the work equally. When it comes to taking meeting minutes and filling out the weekly report, we decided to rotate each week. The team member that is on rotation for the week, has to take notes for our team meetings, meet with our advisor, meet with the spon,sor, and fill out the report. We help each other out by taking our notes for more effectiveness and subbing in when the team member is not able to attend the meeting. Before submitting the weekly report, we all take a good look at it and make sure everything is in good order before submitting.

When it comes to making a report for the sponsor we all work together on a small PowerPoint presentation and assign each other parts to present. Each week we meet and discuss what needs to be done for the project and who can work on what task. The work role changes for each member since we all can’t work on one part together because of our schedules and sometimes other tasks need to be worked on as well. In the end, we all try to work together as best we can, but we make sure everyone is doing good and if someone needs help we ask to see who is available to help.

1. **COMMUNICATION PLAN**:

# 

* + Team: With our team, we communicated through text and Microsoft Teams. We had flexible meetings throughout this project but had regular meetings scheduled on Thu at 4:00 pm
  + Faculty Advisor: We communicated through email but also had regular meetings on team every Tue 4:00 pm
  + Autix: We and the Autix team have a Slack channel dedicated to our communications and we also hold regular updates on Wednesdays at 3:30 pm on Zoom.

1. **TIMETABLE**

During week one of the project, we were introduced to the company that we are working with. We got to meet with the CEO, Jake Hamann, the data scientist, Andrew Bartnik, and Binko. We were introduced to the project in more depth and what the company was trying to achieve through this project. We discussed any ideas we came up with for the project. During week two, we received a training data set from the data scientist. We set up Slack to communicate with each other. In week three, we got access to Azure DevOps, which is a platform used to collaborate. During week four, we all did research and found three methods that we could use to organize and analyze the data. We then presented our findings to the Autix team. We then presented our proposal presentation in class, showing our progress on the project.

1. **EVALUATION**   
     
   We believe our project is successful because we have taken the necessary steps to ensure that our work is approved and well-received by our mentors. We set deadlines and objectives within the first two weeks, making sure we fully understood the project description, objectives, and expected deliverables. We had productive discussions with our mentors, Andrew and Jake, to clarify the tools and data requirements. Once we obtained the required data, we wasted no time and immediately started working on the project. We also kept our work in check by consulting with our faculty advisor, Dr. Chandra, for guidance and support throughout the project. We employed various approaches to address the problem at hand, and one of these solutions received approval from our mentor, Andrew. This milestone marked a significant step forward, and we are now actively engaged in the next phase, which involves testing our model. We received valuable feedback from Jake, the founder of the company, as well as Binko, the CTO of the company. Their positive feedback further validates our project's quality and success. Overall, our project's success is a result of meticulous planning, collaboration with mentors, and a commitment to delivering high-quality work.
2. **CONCLUSION**   
     
   This is a good project because we are putting our coding and analysis skills to work
3. on a real-life data set. This will give us more experience with analyzing unstructured data coming from user input. It allows us to work with complex data that is challenging to clean and work with. At the end of the project, we hope to be able to give a price estimate for a car based on its sentimental value and its make, model, past owners, etc.
4. **CONTACT INFORMATION**

* Mehedi Toufiqe.
  + [mht190001@utdallas.edu](mailto:mht190001@utdallas.edu)
  + <https://www.linkedin.com/in/mehedi096/>
* Sayema Rahman
  + [sfr190001@utdallas.edu](mailto:sfr190001@utdallas.edu)
  + <https://www.linkedin.com/in/sayemarahman>
* Marcos Munoz
  + [mxm200013@utdallas.edu](mailto:mxm200013@utdallas.edu)
  + <https://www.linkedin.com/in/marcos-munoz-9160aa233/>
* Yaseen Mohammed
  + [yam200000@utdallas.edu](mailto:yam200000@utdallas.edu)
  + <https://www.linkedin.com/in/yaseenam/>

1. **RESOURCES & SOURCES**

Resources:

* Python
* ChatGPT
* Google Collab
* Llama
* Slack
* Company’s Internal Datasets.

Sources:

* <https://pypi.org/project/llama-cpp-python>
* <https://platform.openai.com/tokenizer>
* <https://textblob.readthedocs.io/en/dev/>
* <https://textblob.readthedocs.io/en/dev/>

1. **APPENDIX**

Python: https://www.python.org/

ChatGPT: https://openai.com/

Google Colab: https://colab.research.google.com/

Llama: https://pypi.org/project/llama-cpp-python

Slack: https://slack.com/

Classic Cars Data Source: https://www.classiccars.com/

OpenAI Tokenizer: https://platform.openai.com/tokenizer

TextBlob: <https://textblob.readthedocs.io/en/dev/>

1. **Print Name/Signatures/Date**:

| **Sayema Rahman**  *Team Member 1* | **Marcos Munoz**  *Team Member 2* |
| --- | --- |
| **Mehedi Toufiqe**  *Team Member 3* | **Yaseen Mohammed**  *Team Member 4* |
|  |  |
| **Andrew Bartnik**  *Company Mentor Name*, Company Mentor | **Dr. Noirrit Chandra**  *Faculty Advisor Name*, Faculty Advisor |