

Capstone Session - 1

Used Cars Price Prediction Project

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Agenda

- 1) Used Cars Price Prediction Problem Statement
- 2) Solution Steps Walkthrough
- 3) General Best Practices
- 4) Q&A

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Used Cars Price Prediction Dataset

- The dataset consists of **7253 rows and 14 columns**.
- There are **missing values** in the Engine, Power, Seats, New_price, and Price columns.
- Price : The price of the used car in **INR 100,000 (Target Variable)**
- Kilometers_Driven values have an incredibly high range. We should check a few of the extreme values to get a sense of the data.

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

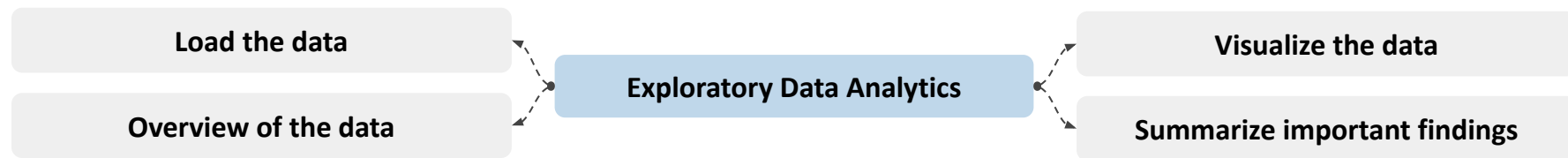
Problem Definition and Objective

- There is a huge demand for used cars in the Indian Market today. As sales of new cars have slowed down in the recent past, the pre-owned car market has continued to grow over the past years and is larger than the new car market now. Cars4U is a budding tech start-up that aims to find footholes in this market.
- In 2018-19, while new car sales were recorded at 3.6 million units, around 4 million second-hand cars were bought and sold. There is a slowdown in new car sales and that could mean that the demand is shifting towards the pre-owned market. In fact, some car owners replace their old cars with pre-owned cars instead of buying new ones. Used cars are very different beasts with huge uncertainty in both pricing and supply. Keeping this in mind, the pricing scheme of these used cars becomes important in order to grow in the market.

Objective:

- Explore and visualize the dataset.
- Build a model to predict the prices of used cars.
- Generate a set of insights and recommendations that will help the business.

Solution Step 1



Example of questions that can be answered by EDA:

1. Is there any missing values in the dataset?
2. Explore the summary statistics of the dataset?
3. Is there any extreme values in the dataset?
4. Is there skewed distribution in the dataset?
5. Do we need to log_transformation?
6. Check whether there is a correlation between dependent and independent variables?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Solution Step 2



Example of questions that can be answered by Data Preprocessing:

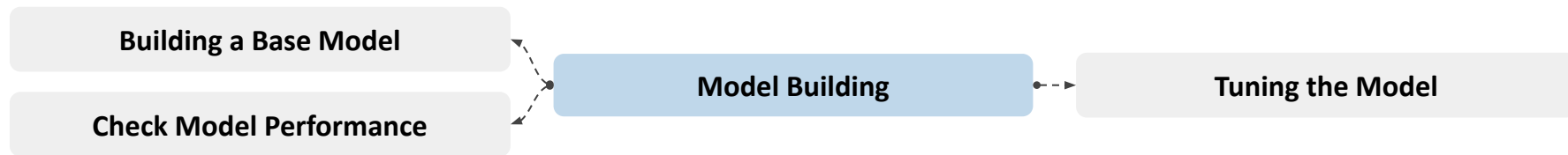
1. Do we need to do feature engineering?
2. Are the missing values needs to be imputed or dropped?
3. Why we should drop 'Name','Price','price_log','Kilometers_Driven' from X before splitting?
4. Do we need to convert categorical variables into dummy variables in the train and test data?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Solution Step 3



Example of questions that can be answered by Model Building:

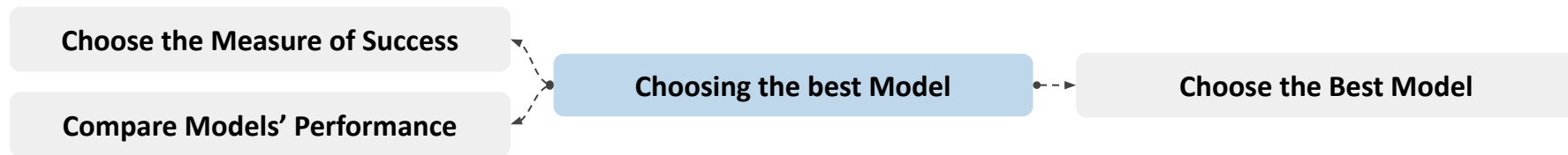
1. Can we use Logistic Regression to solve this problem?
2. Is the model overfitting the training data?
3. Can we improve the model performance by hyper parameter tuning?
4. What are important features?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Solution Step 4



Example of questions that can be answered by Model Selection:

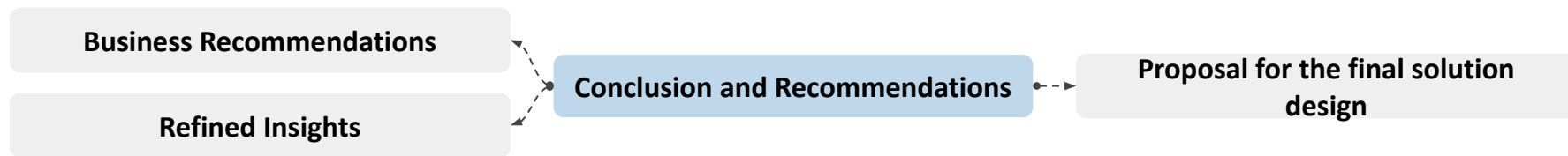
1. What is the metric (Measure of Success) for this business problem?
2. What is the best model?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Solution Step 5



Example of questions that can be answered by Conclusions and Recommendations:

1. What are the refined insights from EDA and model building?
2. What observations and insights can be drawn from the measure of success?
3. How do different techniques perform? Which one is performing relatively better? Is there scope to improve the performance further?
4. What is proposal for final solution design? Why is this the best solution to adopt? What are expected benefits and costs (assume numbers) of this solution design?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

General Best Practices

Some of the best practices for submission:

- **Address all key questions in the rubric:** Make sure to read the rubric carefully and understand all the requirements. Address all the key questions asked in the rubric in your submission.
- **Provide observations and insights:** Provide observation and insight for every important output, such as plots, summary statistics, missing values detection and treatment. This will help to make your work more understandable and actionable.
- **Explain your design steps:** Explain the steps you took to design your solution approach. This will help the reader to understand the overview of your solution approach and how you arrived at your final model.
- **Build different models:** Try to build different algorithms and provide observations for all the techniques. This will help to compare and contrast the performance of each technique and make better decisions.

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

General Best Practices

Some of the best practices for submission:

- **Improve the model by hyperparameter tuning:** Always try to improve the model by hyperparameter tuning. This will help to avoid overfitting and make your model more generalizable.
- **Select the performance metric that best fits the business objectives:** Choose the performance metric that best fits the business objectives. This will help to ensure that your model is relevant and useful to the business.
- **Interpret potential benefits from the model:** Provide an interpretation of potential benefits from the model. This will help the business to determine the next steps and make informed decisions based on your work.

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Q&A

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.