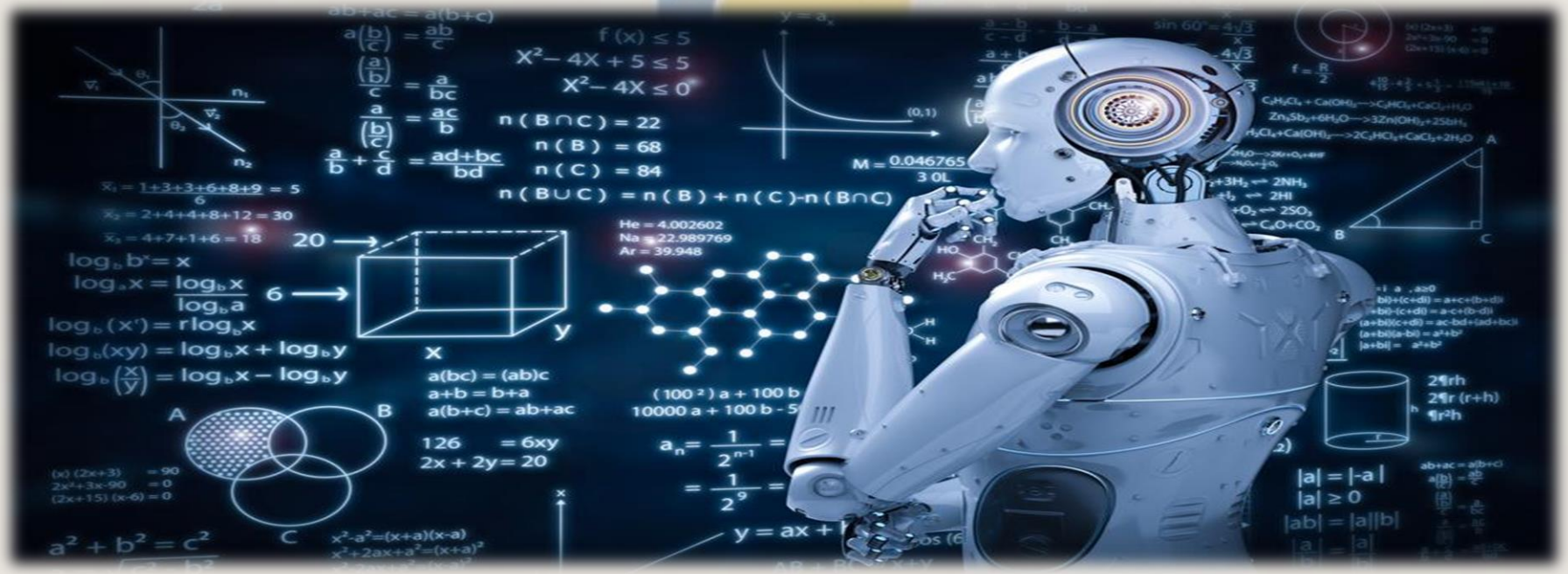
Machine Learning_Model

FOR

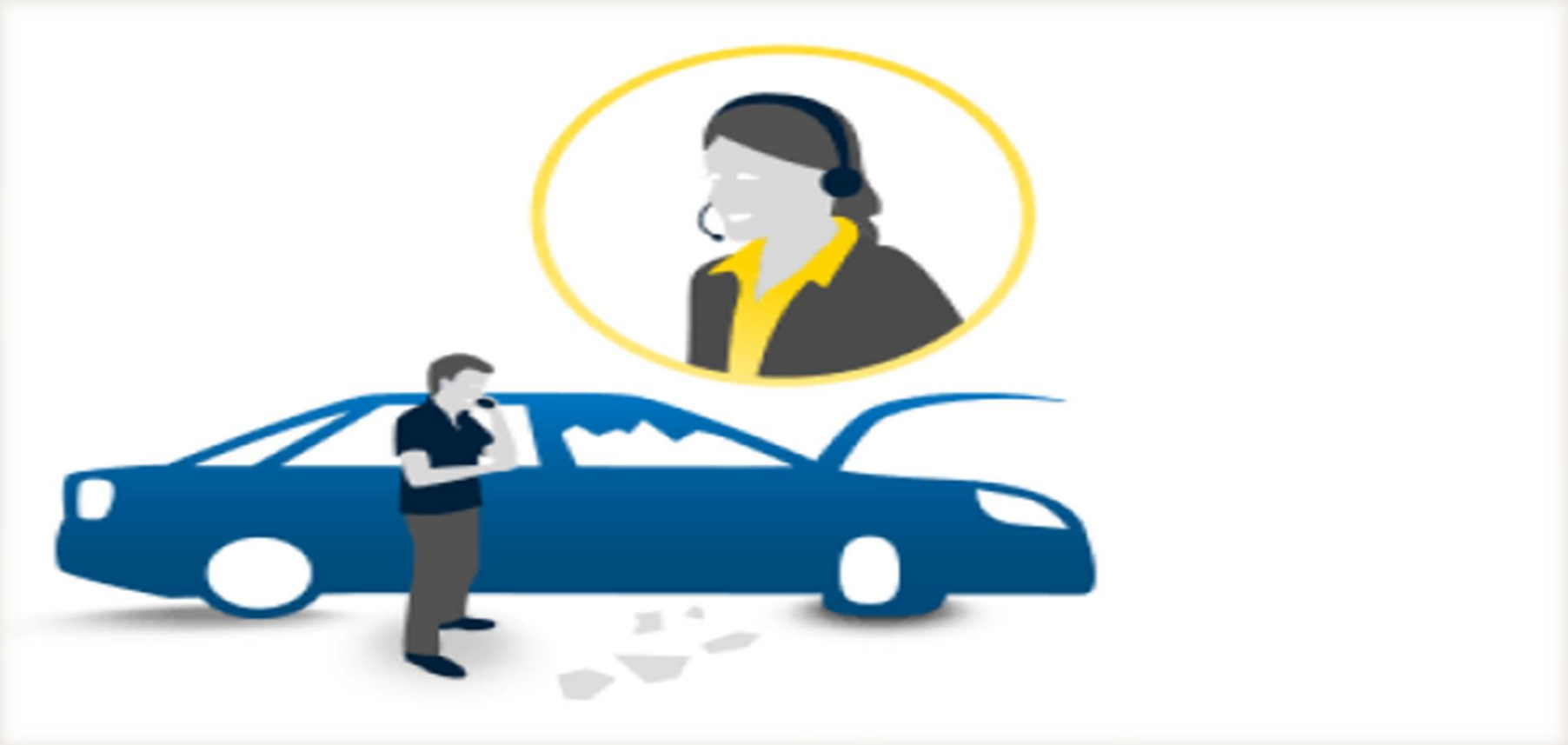
Auto_Insurance_Industry.....



TABLE_OF_CONTENT...

- # AIM OF THE PROJECT

The aim of thr project is to build a Machine Learning Model to perdict whether an owner will initiate an auto insurance claim in the next year.



- # BACKGROUND

The auto insurance industry is witnessing a paradigm shift. Since auto insurance company consists of homogenous good thereby making it difficult to differentiate product A from product B, also companies are fighting a price war (for insurance price). On top of that, the distribution channel is shifting more from traditional insurance brokers to online purchases, which means that the ability for companies to interact through human touchpoints is limited, and customers should be quoted at a reasonable price. A good price quote is one that makes the customer purchase the policy and helps the company to increase the profits.

Also, the insurance premium is calculated based on more than 50+ parameters, which means that traditional business analytics-based algorithms are now limited in their ability to differentiate among customers based on subtle parameters.

- # USE CASE

The model shall mainly support the following use cases:

1. Conquering Market Share: Capture market share by lowering the prices of the premium for the customers, who are least likely to claim.

2. Risk Management: Charge the right premium from the customer, who is likely to claim insurance in the coming year .

3. Smooth Processing: Reduce the complexity of pricing models. Most of the transactions are happening online with larger customer attributes (thanks to the internet and social media). Harness the power of huge data to build complex ML models.

4. Increased Profits: As per industry estimate 1% reduction in the claim can boost profit by 10%. So, through the ML model, we can identify and deny the insurance to the driver who will make a claim. Thus, ensuring reduced claim outgo and increased profit.

- # DATASET DESCRIPTION

The project involves the use of a dataset with 600k training data and 57 features/data. In the train and test data, features that belong to similar groupings are tagged as such in the feature names (e.g., ind, reg, car, calc). In addition, feature names include the postfix bin to indicate binary features and cat to indicate categorical features. Features without these designations are either continuous or ordinal. Values of -1 indicate that the feature was missing from the observation. The target column signifies whether a claim was filed for that policy holder.

• PROCESS FLOW

A. EDA (Exploratory Data Analysis):

Analyze the datasets to summarize their main characteristics (with visual methods). A statistical model can be used, primarily EDA can be used to see what the data can tell us beyond the formal modeling or hypothesis testing task.

__TASK PREFORMED AS A PART OF EDA__

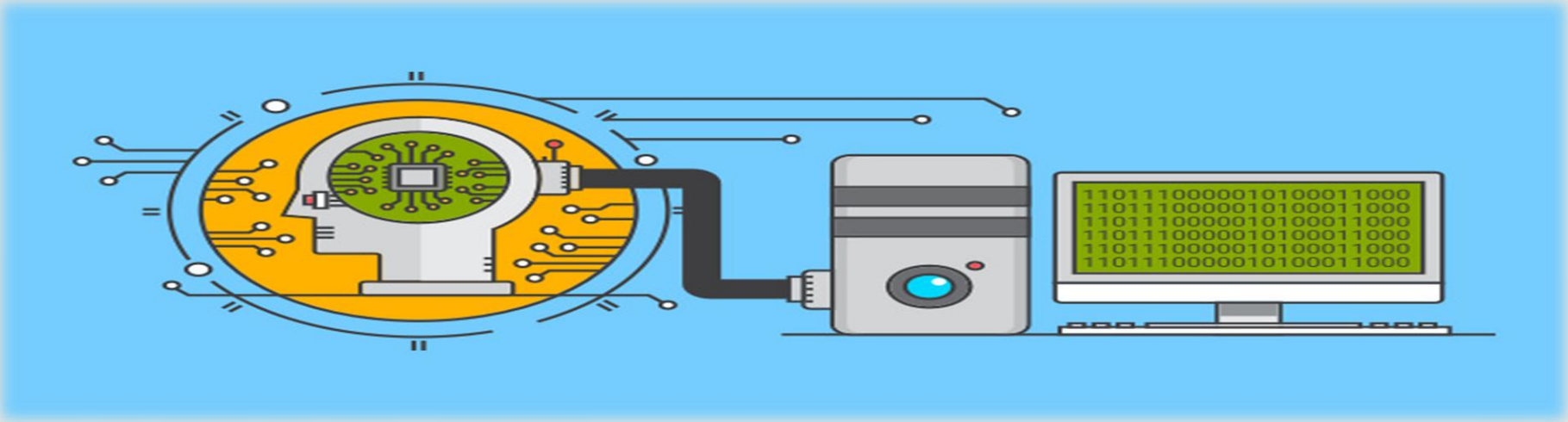
- Shape of the Data-set
- Unique Values
- Missing Values
- Type
- Describe
- Balance Data
- Categorical & Binary features
- Hot Encoding for Categorical Features
- Inference from data based on Interval , Ordinal & Binary variable
- Correlation Matrix
- Ploting Numerical, Categorical & Binary Features
- Correlations between Features & Target

- Whole Data Exploration
- Summarize the learnings of EDA

B. Machine Learning Modeling:

After EDA, the modeling comes into the process. The process of training an ML model involves providing an ML algorithm (that is, the learning algorithm) with training data. The term “ML model” refers to the model artifact that is created by the training proces

The training data must contain the correct answer (target or target attribute). The learning algorithm finds patterns in the training data that maps the input data attributes to the target (the answer that you want to predict), and it outputs an ML model that captures these patterns.



__Task performed as a part of modeling__

- Feature Engineering
- Selection
- Split Into Training and Testing Sets
- Feature Scaling
- Convert y to one-dimensional array (vector)
- Models to Evaluate
- Comparing five different machine learning Cassification models