

Our Data Analysis was divided into three sections.

What is data exploration ? Data exploration is the process of analyzing a dataset to gain insights and generate hypotheses. It involves understanding the data, its properties and its relationships, as well as its context. Data exploration techniques include visualizing the data, summary statistics, correlation analysis, clustering, and more. It helps to identify patterns, trends, and outliers in the data, and to gain a better understanding of what the data is telling us.

What is data preprocessing ? This includes cleaning, transforming, and organizing data so that it can be used in analytics and machine learning models. Data preprocessing involves a variety of techniques such as data normalization, feature scaling, feature selection, and data transformation. The goal of data preprocessing is to create a dataset that is suitable for analysis and can yield more accurate results.

What is feature engineering ? Feature engineering is a process of transforming raw data into features that better represent the underlying data and can be used to improve the accuracy of machine learning models. In general it involves analyzing the relationships between features and the target variable.

For evaluation we used the coefficient of determination (R^2 or r-squared) which is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable.

Some difficulties we encountered was with the feature selection. And that is very important for our prediction. Because some features may be good but combining them may be even better. For example we do have total rooms. And we also have total bedrooms . Yet the ratio of these two might be more intriguing. How many rooms of those rooms are bedrooms. And the second challenge was to convert some categorical data to numerical features in order to make our life easier. For instance, using the categories from the column called ocean_proximity we created five new numerical columns and we dropped the first one.

In essence, all of these algorithms function like a black box. We provide them with certain information as input, and they provide an output. So a huge problem is that there is a tremendous trade-off between accuracy and interpretability. For instance, the linear regression model is simple to understand yet frequently has poor accuracy. Artificial Neural Networks, on the other hand, have a high degree of accuracy but are challenging to understand, making people less inclined to believe the results they provide.

For that specific reason we came up with the idea to use Explainable AI.

What is Explainable AI ? Explainable AI (XAI) is an area of artificial intelligence (AI) that focuses on creating AI systems that can explain their decisions and behaviors. XAI seeks to create AI systems that are more transparent, interpretable, and accountable. This is important for ensuring that AI systems are used responsibly and ethically. XAI techniques such as feature importance, partial dependence plots, and counterfactual explanations can help to explain the decisions made by AI systems and provide insights into how they work.

Explainable AI can be used in a variety of applications such as debug AI systems, or to identify potential bias in AI systems. Additionally, XAI techniques can be used to help users better understand and interact with AI systems.