**PROJECT PROPOSAL**

***Siddhesh Kadam***

***Net Id – vz3642***

***Capstone Project. -Fall 2023***

**Machine Learning Analysis of California Housing Data**

**Introduction:**

* This research project aims to implement machine learning algorithms on the California housing dataset to gain valuable insights into the housing market in California.
* The dataset provides an accessible resource for teaching the basics of machine learning. By conducting data cleaning and preprocessing, the relationships between various variables will be explored, and predictive models for median house values in California districts will be developed.
* Additionally, this project will leverage AWS services, specifically Amazon Sage Maker, to enhance the data analysis process.

**Objectives:**

* Data Cleaning and Preprocessing: Perform necessary steps to clean the dataset, handle missing values, and ensure data quality for accurate analysis.
* Exploratory Data Analysis: Conduct a comprehensive analysis of the dataset to identify patterns, correlations, and outliers, providing insights into the California housing market.
* Feature Engineering: Apply advanced techniques to generate new features or transform existing ones, aiming to improve the predictive power of the machine learning models.
* Predictive Modeling: Utilize machine learning algorithms, including regression models, ensemble methods, and deep learning models, to develop accurate predictive models for median house values.
* Evaluation and Visualization: Evaluate the performance of the developed models using appropriate evaluation metrics and visualize the results in an interpretable manner, aiding in understanding the housing market dynamics.
* Data Wrangling and Analysis: Utilize AWS services, particularly Amazon Sage Maker, for efficient data wrangling, cleaning, manipulation, and in-depth analysis of the California housing dataset.
* Data Visualization and Dashboard Creation: Leverage AWS services, including Amazon Quick Sight, to create visually appealing and interactive dashboards, presenting key findings from the analysis and facilitating data-driven decision-making.

**Conclusion:**

**By successfully achieving these objectives, this research project aims to enhance the understanding and application of machine learning techniques while providing valuable insights for housing market analysis in California. We will create a live interactive Tableau Dashboard as well for the final output. The utilization of AWS services, will enable efficient data processing, modeling, and visualization, contributing to the overall knowledge in this field.**

*I am excited to undertake this project under the guidance of my project supervisor,* ***Prof. James Tandon.*** *The outcomes of this research will not only enrich my profile but also make a significant impact in the field of machine learning and housing market analysis.*