# Amar Chaurasiya Data Scientist

**■** amarchaurasiya124@gmail.com

7039190879

Andheri, Mumbai

www.linkedin.com/in/amar-chaurasiya163788262

## **Projects**

#### **Car Price Prediction Dataset**

- Created a predictive model to estimate car prices based on body style, fuel type, horsepower, engine size, and other vehicle characteristics.
- Size of the dataset: 205 x 26.
- As part of the data preparation process, I handled missing values, encoded categorical variables, and scaled numerical features to improve model performance.
- Used a number of machine learning methods, such as **support vector machine** (**grid search score**: 0.7465) and **linear regression** (**RMSE**: 2377.51, **R**<sup>2</sup>: 84.65%).

## **Housing Sales Prediction**

- Created a predictive model to estimate housing costs utilizing features such as square footage, number of rooms, location, and other property-related parameters.
- Size of DataSet: 20640 x 10.
- To improve model performance, extensive data preparation was carried out, including missing value management, categorical variable encoding, and numerical feature scaling.
- Explored and contrasted multiple regression techniques, such as **linear regression** and **support vector machine**, and achieved model accuracy of 95% on the test data.

## **Heart Disease prediction**

- Developed a heart disease prediction model using a dataset with features such as age, sex, chest pain type, blood pressure, cholesterol, EKG results, and more.
- Size of DataSet: 270 x 14.
- The project involved comprehensive exploratory data analysis (EDA) and data preprocessing, including label encoding to enhance model performance.
- Implemented multiple algorithms, including **logistic regression** (accuracy 81.48%), **decision tree** (accuracy 68%), and **support vector machine** (SVM), which achieved the highest accuracy of 82.59% using **grid search cross-validation**.

## **Term Deposits**

- Developed a model using the term deposit of the bank dataset. The key features included age, marital status, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, and outcome.
- Size of DataSets: 4521 x 17.
- I performed exploratory data analysis (EDA) and applied label encoding to the data. for model development.
- Implemented and evaluated several models, including **logistic regression** (accuracy: 90.19%), **support vector machine** (SVM) with hyperparameter tuning (**grid search cross-validation** score: 89.49%), and **XGBoost** (**grid search cross-validation score**: 89.94%).

## Digit Classification using Convolutional Neural Network (CNN)

- Developed a CNN model to classify handwritten digits from the MNIST dataset.
- MNIST is an actual dataset built consisting of 60,000 in training & 10,000 images for testing.
- The design model consists of convolutional & pooling layers followed by multiple ANN layers.

## Sentiment Analysis of Amazon Reviews (NLP)

- Conducted sentiment analysis on 10,000 Amazon reviews using the **VADER sentiment analysis** library to classify reviews as positive or negative.
- Abstracted insights into customer sentiment trends enhance understanding of product feedback through efficient text processing and scoring techniques.

## Finance Report Summary Dashboard

- The dashboard, designed using Power BI, created an interactive dashboard to monitor financial KPIs, including gross sales (111M), total sales (103M), profit (15M), and units sold (972K).
- Integrated dynamic date-based filtering to analyze sales trends and profit variations over time.
- Visualized regional performance through a profit-by-country map, providing insights into geographical contributions
- Represent sales by product and segment using bar charts, pie charts, and line charts for comprehensive analysis.

#### HR Analytics Dashboard

- The dashboard, established using Power BI, highlights key metrics: 961 employees, a 13.8% attrition rate (133 employees), an average age of 37, a salary of \$6.3K, and a tenure of 6.9 years.
- Attrition is highest among males (88), employees aged 26-35, and in life sciences (44%) and medicine (35%). Roles most affected include lab technicians (62) and research scientists (47).
- Attrition is linked to salaries below \$5K and low job satisfaction. Interactive buttons for HR, R&D, and Sales provide detailed department insights to aid retention strategies.

### **Skills**

Python	SQL
Excel	Machine Learning
Power BI	Tableau
Deep Learning	Natural Language Processing (NLP)
Statistics	Gen AI

### Certificates

Data Science & Artificial Intelligence	Basic, Intermediate to Advanced SQL 🛭
(From Boston Institute of Analytics) (29-06-24/15-12-24)	Probability and Statistics using Python ☑

## Google Colab Python 🛮

## **Education**

2021 – 2024 Mumbai, India	Bachelor of Science in Information Technology Tolani College CGPA 7.08
2019 – 2021 Mumbai, India	HSC Tolani College Percentage: 77.78%
2018 – 2019 Mumbai, India	SSC S.T. Arnold's High School Percentage 76.77%