# Furkan Ismail Mistry

Data Scientist and Machine Learning Engineer

Mumbai | 7303307368 | furkanmistry47@gmail.com

## Profiles

Furkan Mistry
Linkedin

FurkanMistry
Github

fmistry42
HackerRank

## Summary

Versatile Data Scientist and Machine Learning Engineer with a passion for acquiring hands-on industry-level experience in data analytics and machine learning. Currently pursuing a Master's degree in Big Data Analytics with a focus on mastering skills in Machine Learning, SQL, MongoDB, and Deep Learning. Dedicated to converting academic learnings into practical expertise, seeking opportunities to contribute to cutting-edge projects.

## Experience

**The Tann Mann Foundation** — **February 2022 - May 2022**
Machine Learning Yolo intern — Remote

As an intern, I had the opportunity to work on an exciting computer vision project where I was responsible for labeling images of different objects using LabelImg, building models, and testing their accuracy.

My task involved using LabelImg to manually label images of various objects. I then used the labeled images to train different models, such as Darknet YOLOv3, and tested their accuracy on a separate set of images.

## Education

**Kishinchand Chellaram College, Mumbai** — **March 2020 - April 2023**
Information technology — Bachelor of Science
9.77

**St Xaviers College, Mumbai** — **July 2023 - Present**
Big Data Analytics — Masters in Science

## Certifications

**Machine Learning** — **September 2021**
NPTEL

**Natural Language Processing with Classification and Vector Spaces** — **Nobember 2023 - Present**
Coursera - DeepLearning.AI

## Skills

**Python Libraries - Pandas, Numpy, scikit learn, tensorflow, keras, matplotlib, seaborn**
●●●●○
Pandas

**SQL**
●●●●○

**Machine Leaning**
●●●●○

**NoSQL**
Firebase
●●●●○

**Deep Learing**
●●●●○

**Feature Engineering**
●●●●○

**Data Visualisation and Analysis**
PowerBI
●●●●○

**EDA**
●●●●○

## Projects

### Financial News Sentiment Analysis

https://colab.research.google.com/drive/1giAkVuSycjCENlzz5FRM_lIxxrRW6gCt#scrollTo=ibmAtVenehK2

Developed and implemented a Sentiment Analysis System for financial news articles using Python, NLTK, and machine learning models such as Linear SVC, Logistic Regression, Multinomial Naive Bayes, and Bernoulli Naive Bayes. Collected and preprocessed data from reputable financial news sources, performed model evaluation, and conducted comparative analysis. Selected Linear SVC, achieving an accuracy of 83.81%, as the dataset exhibited linear separability. Integrated real-time prediction for live sentiment analysis of new articles and utilized web scraping for live testing on a website. Extracted linguistic characteristics from articles, visualized model accuracy, and recommended enhancements. Acknowledged project limitations and maintained version control using Google Colab. Recognized for practical applicability with live website testing.

### Resume and JD Similarity Score API

https://github.com/FurkanMistry/resume-similarity-scoring

Developed a dynamic API using Flask in Python for efficient resume screening. The API accepts job description details, including job description, experience, and required skills, and processes uploaded resumes. Implemented comprehensive text processing techniques, including Jaccard similarity and other similarity metrics, to clean and tokenize the resume text. Incorporated libraries such as docx2txt, PyPDF2, and NLTK to ensure compatibility with both Word and PDF formats. Leveraged the CountVectorizer and various similarity metrics, including Jaccard similarity, to assess the similarity between the job description and the cleaned resume text. Demonstrated optimal performance with the cosine similarity method, providing accurate matching scores. This API serves as a powerful tool for automating the resume screening process, enhancing the speed and accuracy of candidate selection.

### Fraud Detection

Developed a fraud detection system leveraging advanced machine learning techniques for credit card transactions. The project involved benchmarking the Isolation Forest Algorithm and Local Outlier Factor on a highly unbalanced dataset containing 492 frauds out of 284,807 transactions in September 2013. Implemented anomaly detection due to the limited instances of fraudulent data. Utilized Isolation Forest's unique approach of isolating anomalies and achieved a remarkable 99.74% accuracy, outperforming the Local Outlier Factor algorithm. Successfully detected 73 errors with Isolation Forest compared to 97 errors with LOF. The project highlights a strategic use of machine learning to combat dynamic fraud patterns in financial transactions.

### Image Classification using CNN

https://github.com/FurkanMistry/ImageClassificationCNN

Implemented an image classification project using Convolutional Neural Networks (CNN) for the CIFAR-10 dataset. Utilized TensorFlow and Keras to develop the model architecture. Loaded and preprocessed the dataset, consisting of 60,000 32x32 color images in 10 different classes. Trained a neural network with both a dense (fully connected) layer and a CNN layer for comparison. The dense layer achieved an accuracy of approximately 49%, while the CNN achieved a higher accuracy of about 70% on the test set. Employed metrics such as precision, recall, and F1-score to evaluate the model's performance on individual classes. Visualized predictions on sample images, demonstrating the model's ability to classify objects like airplanes, automobiles, birds, cats, and more. This project serves as a foundational exploration into image classification using deep learning techniques.