# Data Science Portfolio #

## End to End Projects ##

### [Alzhimers CV-BOLD Classification]( https://github.com/ZahraSangboriToroghi1/Machine-Learning/tree/main/Classification/Alzhimers%20CV-BOLD%20Classification) ###

\* Trained multiple classifiers that improved the classification performance of imbalanced Alzheimer’s fMRI images by more than 12% compared to state-of-the-art on the same data.

\* Analyzed, visualized, and discussed the results with a team of neurological researchers to have a better understanding of the results and Alzheimer’s disease.

\* Analyzed, visualized, and reported the results and submitted a research paper to the ISPr 2023 scientific conference.

### \_\_[Real Time Sign Language Interpretation Web Application]( https://github.com/ZahraSangboriToroghi1/Computer-Vision/tree/main/Real%20Time%20Sign%20Language%20Interpretation%20App)\_\_ ###

\* Developed a real-time sign language interpretation application using React.js, TensorFlow, and tensorflow.js and deployed it on IBM cloud servers.

### [Building Movie Recommendation System using Pyspark]() ###

\* Building a recommendation engine using Alternating Least Squares in PySpark and using the popular MovieLens dataset and the Million Songs dataset.

### [Real Time Car Plate Detection Mobile Application]() ###

\* Building a real-time car plate detection mobile application using TensorFlow and EasyOCR.

----

# Skill Based Projects #

## Generative AI ##

### Fine Tuning LLMs ###

\* [Finetune Falcon-7b with LoRA]():

\* [Instruction Fine Tuning T5 LLM for Summarization]()

### Reterival Augmented Generation (RAG) Application ###

\* [RAG using Gemma & Faiss Vector DB]()

### Prompt Engineering ###

## Machine Learning:

### Regression

\* \_\_[Automobile price prediction]( https://github.com/ZahraSangboriToroghi1/Machine-Learning/tree/main/Regression/Automobile%20price%20prediction)\_\_: Utilize Python to implement end to end data science pipeline to predict the price of old Automobile based on the given features.

### Classification

\* \_\_[Sensor Activity Recogniation]( https://github.com/ZahraSangboriToroghi1/Machine-Learning/tree/main/Classification/Sensor-activity-recognition/codes)\_\_: Classifying the output of eight sensors into five activities and studied the effect of changing window

sizes and axel combination.

\* \_\_[Alzhimers CV-BOLD Classification]( https://github.com/ZahraSangboriToroghi1/Machine-Learning/tree/main/Classification/Alzhimers%20CV-BOLD%20Classification)\_\_: Utilized Python to develop supervised machine learning techniques to classify imbalanced Alzheimer’s CVBOLD data, which enhanced the classification performance by 10%.

### Clustering

\* \_\_[Find the best location to open a new Gym]( https://github.com/ZahraSangboriToroghi1/Machine-Learning/tree/main/Clustering/Finding-the-best-Tornoto-neighborhood-to-open-a-new-gym)\_\_: Utilized python to implement unsupervised techniques to helping the business owner to increase his revenue by finding the best neighborhood to open a new gym.

\* \_\_[Customer identification for mail order products]( https://github.com/ZahraSangboriToroghi1/Machine-Learning/tree/main/Clustering/Customer%20identification%20for%20mail%20order%20products)\_\_: Utilized python to implement unsupervised techniques to helping the business owner to increase his revenue by finding the best neighborhood to open a new gym.

---

## Deep Learning

### Classification

\* \_\_[Melenoma Classification]( https://github.com/ZahraSangboriToroghi1/Deep-Learning)\_\_: Classifying malignant Melanoma using skin lesion images using CNN-based classifiers.

---

### Computer Vision

\* \_\_[Object Tracking using Particle Filter]( hub.com/ZahraSangboriToroghi1/Computer-Vision/tree/main)\_\_: Implemented particle filter to track walking object in video

\* \_\_[Pose Estimation and Squat counter]( https://github.com/ZahraSangboriToroghi1/Computer-Vision/tree/main/Pose%20Estimation%20%26%20Squat%20Counter)\_\_: Utilize python to develop a real-time pose estimation and squat counter using MovingNet lightning.

\* \_\_[Object Detection Deployed on FastAPI]( https://github.com/ZahraSangboriToroghi1/Computer-Vision/tree/main/Real%20Time%20Sign%20Language%20Interpretation%20App)\_\_:

---

### Natural Language Processing

\* \_\_[Sentiment Analysis web app]( https://github.com/ZahraSangboriToroghi1/Natural\_Language/tree/main/Sentiment-analysis)\_\_: Web application for classification of reviews, using deep learning model implemented in PyTorch and deployed on Amazon SageMaker.

\* \_\_[Plagirasm Detector web app]( https://github.com/ZahraSangboriToroghi1/Natural\_Language/tree/main/plagiarism-detector-web-app)\_\_: Creating plagiarism detector trained on LSC and containments features and deployed on AWS SageMaker.

\* \_\_[Data Science Resume Selector]( https://github.com/ZahraSangboriToroghi1/Natural\_Language/tree/main/Data-Science-Resume-Selector)\_\_: Selecting the resume that are eligbile to data scientist postions, the dataset used contains 125 resumes, in the resumetext column. Resumes were queried from Indeed.

---

### Time-series Analysis

\* \_\_[Power consumption prediction]( https://github.com/ZahraSangboriToroghi1/time-series)\_\_: Real time prediction for power consumption using DeepAR on AWS.

---

### Data Visulization:

\* \_\_[Immigrants to Canada data visualization]( https://github.com/ZahraSangboriToroghi1/Data-Visualization)\_\_: Visualizing the data of the immigrants to Canada using different visualizing libraries in Python.

\* \_\_[Geospatial visualization of San Francisco Police Department Incidents]( https://github.com/ZahraSangboriToroghi1/Data-Visualization)\_\_: Visualizaing the geospatial data of the San Francisco police department incidents for the year 2016.

---

### Spark

\* \_\_[San Diego Rainforest Fire Predicition]( https://github.com/ZahraSangboriToroghi1/S-P/tree/main/Cluster%20Analysis%20of%20the%20San%20Diego%20Weather%20DataPredicting the occurrence of rainforest fire in san Diego using weather data collected by san Diego weather center.

\* \_\_[Cluster Analysis of the San Diego Weather Data]( https://github.com/ZahraSangboriToroghi1/S-P/tree/main/San%20Diego%20Rainforest%20Fire%20Predicition)\_\_: Utilizing pyspark to implement unsupervised learning model to cluster the san Diego weather data so as to better understand the occurrence of the rainforest fire.

---

### Data Modeling

\* \_\_[Songs App data modeling using Apache Casandra]()\_\_: Create an Apache Cassandra database which can create queries on song play data to answer analysis questions.

---