FAISAL IBRAHIM

PORTFOLIO

Bradley Stoke, Bristol, United Kingdom, BS32 8EH ◆ 07425520098 ◆ faisal.ibrahim.work@gmail.com

LinkedIn: https://www.linkedin.com/in/faisal-ibrahim-611b98256
GitHub: https://github.com/FaisalBu123?tab=repositories

EXTRA EMPLOYMENT DETAILS

Machine Learning Engineer Internship, 06/2024 – 09/2024
University of the West of England - Bristol

- File 'Internship - Pose Detection' in the link below is just some of the work done so far (informally written)
- FaisalBu123/Pig-Emotion-Detection-Machine-Learning: Partnering with SRUC (Scotland's Rural College) to analyse videos of pigs for emotion analysis using pose detection. Utilising YOLOv8 and transfer learning to train deep learning models to track the pose changes of the pig. Enabling effective animal welfare monitoring.

Machine Learning/Software Engineer Internship 02/2024 - 05/2024
Think Pacific - Bristol

- Analysed the RAVDESS dataset that consisted of videos of a person speaking in different emotions. The pitch, sentiment, and words said were analysed.
- Logistic Regression model was trained to make predictions of the sentiment of the words said.
- FaisalBu123/Internship-emotion-detection-software-EDS-using-machine-learning: Created an emotion detection software (EDS) using machine learning. This project was implemented in schools in Fiji and aided children with deficiencies in emotion detection and thus regulated the training of their social skills. Used Python, OpenCV, Librosa, TensorFlow, and NLTK. Used RAVDESS dataset.

Aviation Engineer, 01/2023 - 01/2023
General Electric – Virtual

- Through this program on Forage.com, I completed practical tasks in energy, power generation, healthcare, and aviation systems.

Credit Card Specialist, 12/2022 - 12/2022
Visa - Virtual

- Through this Token Service Technology program on Forage.com, I completed practical tasks in card authentication and selecting visa API for use cases.

Cyber Security Engineer, 12/2022 - 12/2022
Mastercard - Virtual

- Through this program on Forage.com, I completed practical tasks in designing a phishing email simulation and in interpreting phishing simulation results.

Software Engineer, 11/2022 - 11/2022
Align Technology - Virtual

- Through this program on Forage.com, I completed practical tasks in project planning for a front-end prototype.

OTHER PROJECTS

Third year bachelor's dissertation project-
Investigating the feasibility of a tool that automates the process of detecting skin cancer (malignant melanoma) using traditional computer vision techniques and the ABCDE rule.

- A project like this is nowadays usually done using convolutional neural networks (CNN) which are a subset of machine learning/deep learning. CNNs are like a "black box", which means their decision-making is quite challenging to understand by non-experts in CNNs, such as doctors.
- However, with my project, I simply taught the computer to look for certain features, according to the ABCDE rule of detecting skin cancer, and then quantified the features i.e. gave them numbers, for easier diagnosis. For example, for the 'A' part of the ABCDE rule, which stands for asymmetry, a value between 0-10 would be given, 0 being symmetrical and 10 being asymmetrical. A total score was then derived. This is extremely easy to understand by a dermatologist, GP, or even the general public, as all they must do is just look at a few numbers and then move on to the next feature. I used Python, PyCharm IDE, and OpenCV software library.
- The aim of this project was to investigate whether quantifying the features like this is more useful to a doctor than just a system using deep learning that says whether a skin lesion is just benign or malignant. Many doctors complain that systems like this are hard to understand. Many details were included in my proposed system that made sure the doctors feel secure about using the system without being intimidated by the artificial intelligence used (for details read 'Dissertation -23-24' document).
- The system could also be used at home by the public, for example if human contact is not an option such as if the patient has Covid-19 or if the nearest GP is far away and the patient wants to know how serious it is before making the trip.
- The project analysed a large dataset containing thousands of images. The data were cleaned, resized for consistency, cropped, etc. The project involved significant technical research and numerical analysis as well. A written professional report was submitted of the findings and some recommendations. The project improved my Python programming, research, data analysis, communication, time management, project management, computer vision, and Excel skills.
- The project was submitted on time, and earned a first-class grade, both for the presentation and the written report. The project was successful and after conducting a survey, the answers from both medical professionals and the general public indicated that a vast majority of them would use it. The medical professionals liked the fact that the project aims to assist rather than replace them, which was one of the most important goals when designing the system.
- I am now proficient in image processing and computer vision tasks and have implemented various computer vision algorithms using OpenCV libraries. The task included image processing techniques such as filtering, edge detection, gamma correction, Otsu Thresholding, segmentation, computing the circularity and compactness, measuring the diameter using minimum enclosing circle, colour detection using masks, and feature extraction using OpenCV.
- File 'Dissertation -23-24' in the link below shows the dissertation.
- FaisalBu123/Assisting-skin-cancer-detection-using-computer-vision: Investigating the feasibility of a tool that automates the process of detecting skin cancer (malignant melanoma) using traditional computer vision techniques and the ABCDE rule.

Machine learning project-
Predicting medical insurance costs using 8 different regression models and comparing them

- Followed the machine learning pipeline using steps such as exploratory data analysis, data cleaning, and feature engineering. Feature engineering included feature reduction with techniques like PCA and decision tree feature importances. Decision trees have built in 'feature_importances' calculations that were used to rank the features by importance, allowing me to pick only the most important ones. Furthermore, carried out feature scaling and hyperparameter tuning.
- Analysed variable relationships
- Utilised supervised learning.
- Models implemented were – linear regression, linear regression with gradient descent, polynomial regression, random forrest, gradient boost, XGBoost, decision tree, and Multi-Layer Perceptron.
- Models compared using R^2 scores, MSE scores, speed, and explainability.
- Carefully considered the ethics of each stage of the pipeline
- Implemented different visualisations and delivered insights to stakeholders
- Predicting-Medical-Cost-Insurance-Machine-Learning-Project/Faisal Abdul-Fattah Medical Insurance Costs.ipynb at main · FaisalBu123/Predicting-Medical-Cost-Insurance-Machine-Learning-Project

Machine learning project-
Classifying music genre using 11 different classification models and comparing them

- Two datasets were analysed, one with 1,000 30 second samples and one with 10,000 3 second samples. The performance of both was compared for all the models
- Fixed problems with overfitting of the smaller dataset using pruning, varying the regularization strength C, dropout layers, etc.
- Implemented novel methods such as a hybrid ensemble model combining a pre trained image based RESNET model (trained on the same song samples but with the image representations) and a random forrest model (trained on the same song samples but tabular representations). This was done to see if the scores of the overfitting models could be improved this way and also for the sake of curiosity. It worked; the scores improved.
- Utilised supervised learning using the Kaggle 'GTZAN Dataset – Music Genre Classification'
- Models implemented were – logistic regression, decision trees, KNN, random forrest, adaboost, XGBoost, SVM, Voting, ANN, CNN, and the hybrid model.
- Models were compared using train accuracy, test accuracy, precision, recall, f1, AUC, speed, and explainability.
- Implemented different visualisations and delivered insights to stakeholders
- Encoding categorical features appropriately depending on their nominal or ordinal type
- Hyperparameter optimisation using grid search or random search depending on the specific case
- Classifying-Music-Genre-Machine-Learning/Faisal Abdul-Fattah Classifying Music Genre.zip at main · FaisalBu123/Classifying-Music-Genre-Machine-Learning

Unsupervised Learning side project-
Using K-means clustering to categorise countries according to their development

- Side project to better educate myself about the social and economic situation of different countries
- This project aims to categorise countries according to their development using features such as child mortality and health spending.
- Classifying countries as developing, developed, or underdeveloped is useful to organisations like HELP International, a non-governmental organisation that helps fight poverty and supplying people of certain countries with relief during natural disasters.
- This project will show which countries are in most need of the organisation's help to allow more effective resource allocation.
- Using the elbow method and silhouette coefficient to evaluate the clustering
- FaisalBu123/Unsupervised-Learning-country-categorisation: Using K-means clustering to categorise countries according to their development.

Machine Learning TensorFlow side project-
Predicting heart disease using CNN and some common risk factors

- Side project done since I am at risk of heart disease genetically, due to widespread heart disease across my relatives. So, it is interesting to analyse heart disease and its risk factors
- Used SMOTE to balance the dataset, different types of encoding, ranked feature importances, etc.
- FaisalBu123/Predicting-Heart-Disease: Predicting heart disease using CNN and some common risk factors

Machine Learning Neural Network side project-
Analysing three different datasets to try to predict the winner of the 2018 World Cup

- Fun casual side project done out of curiosity and my love for football
- Aim was to find out if I could correctly predict the winner (France)
- Used Neural Network and TensorFlow
- Filtering the data, replacing values, creating and combining features, merging columns, etc.
- FaisalBu123/Predicting-2018-World-Cup-Winner: Analysing three different datasets to try to predict the winner of the 2018 World Cup. Aim was to find out if I could correctly predict the winner (France). Used Neural Network and TensorFlow.

Reinforcement Learning Optimisation side project-
Multi-Armed Bandit problem solved using Upper Confidence Bound reinforcement learning algorithm

- Fun casual side project done to increase my knowledge of reinforcement learning and optimisation problems
- This dataset contains rows where each row corresponds to a unique user and each column corresponds to a specific ad. If the value is 1, then the user clicked on the ad, if 0, then they did not.
- The aim is to find which ad is the most effective at getting clicks
- Problem aims for an exploration and exploitation trade off
- Optimisation Objective: Find the ad selection strategy that maximises the total number of clicks (or rewards) over time
- FaisalBu123/Multi-Armed-Bandit-Optimisation-Reinforcement-Learning-: Ad clicking dataset used. This notebook will use the Upper Confidence Bound (UCB) reinforcement learning algorithm.

Search and Optimisation project-
Using 8 different algorithms to solve the Traveling Salesman Problem (TSP) and then comparing the algorithms

- Aim was to find the shortest route that visits each city in a given set exactly once, before returning to the initial city
- Tested 7 Single-Solution-Driven Search Algorithms and 1 Population-Driven Search Algorithm
- The algorithms were: Simple hill climbing, random restart hill climbing, stochastic hill climbing, steepest hill climbing, iterated local search, simulated annealing, tabu search, and finally a genetic algorithm.
- Compared the cost/distance, the efficiency, space and time complexity (using Big O Notation), computational complexity, the convergence speed, the scalability, and the robustness, using visualisations and t-tests whenever possible
- Implemented many improvements to the algorithms to optimise them, and utilised hyperparameter optimisation using grid search or random search.
- Discussed ethical issues around licensing/code/datasets used and how these issues would compare to commercial use
- Presented findings in a jupyter notebook, report, as well as a video presentation (all details shown below)
- **FaisalBu123/Traveling-Salesman-Problem-Search-and-Optimisation: Using 8 different algorithms to solve the Traveling Salesman Problem (TSP) and comparing the algorithms**

Data visualisation/social research side project-
Analysing a group of students' opinions on AI

- Fun casual side project done out of curiosity
- Aim was to find out how much students know about AI, how they learned about AI, how and where they think AI will be used in the future, among other things, and then provide insights.
- Used various visualisations such as violin plot and pie charts
- As a student studying Artificial Intelligence, I find it fascinating to explore the diverse opinions others have about AI. It is important to understand what people think of AI, as learning from other perspectives allows me to understand the social and ethical responsibilities that accompany AI development.
- By identifying people's concerns and fears, we can work towards creating AI technologies that are trustworthy, less intimidating, and reflective of societal values.
- FaisalBu123/Analysing-a-group-of-students-opinions-on-AI: Fun casual side project done out of curiosity. Aim was to find out how much students know about AI, how they learned about AI, how and where they think AI will be used in the future, among other things.

Graph traversal algorithms-
Brute force search and depth first search algorithms used to solve 3 different problems

- The problems solved were the knapsack, 8-puzzle, and N-queens problems
- The aim was to gain a better understanding of how both algorithms work
- FaisalBu123/Breadth-first-and-brute-force-search: Brute force search and depth first search algorithms used to solve 3 different problems. The problems solved were the knapsack, 8-puzzle, and N-queens problems

SKILLS IN DETAIL

Data analysis Data ETL. Visualising data. Identify trends. Descriptive analytics. Providing high quality technical reports of the findings of the data analysis by using R Markdown. Data engineering, data mining, handling structured and unstructured data. Effectively planning and executing analysis workflows to ensure accurate findings, correlations, and insights. Implementing digital solutions tailored to a client's or project's needs that help optimise quality. Perform statistical analysis. Using Python, C and R languages. Using Excel and PowerBI.

Control Systems Design Control Mathematics. MATLAB and Simulink simulations for control system design and analysis. Modeling multivariable control systems. Optimising control systems. Designing and analysing complex modern dynamic system models. Achieved one of the highest module grades (90%).

Robotics – Developing the robot car, prototyping it, direct Vehicle Control Level (VCU) development. Using C++. Integrating actuators and control systems, testing initial versions of robotics systems, communication protocols, optimising code.

Verification & Validation: Ensuring accurate sensor data transmission (via CAN bus) and optimal system performance for the UWE Formula Society. Testing hardware & software (Python/C++). Identifying faults & debugging electronic systems.

Machine Learning: Predictive models to anticipate future needs, detect outliers, optimise performance, etc. Considering ethics. Non-linear programming. Optimisation mathematics. Experienced in executing the machine learning pipeline. Exploratory data analysis. Feature engineering. PCA. Hyperparameter optimisation. Experienced in regression and classification models. Using Jupyter Notebooks, Numpy, Pandas, and SciKit Learn.

Motorsport skills: Analysing data from the UWE Formula Society, such as fuel and speed, to optimise performance of the car (using SQL, Python, and PowerBI). Experienced with Spark. Using version control (Git) and Agile development practices to structure our work. 3D LiDAR Sensing technology, ROS, etc.

Statistical analysis Understanding of statistical analysis plans and reproducible research. Interpreting descriptive statistics and graphics (one and two dimensions). Primary and secondary research. Numerical analysis. Using simulations to calculate confidence intervals.

Data science Handling large amounts of data, cleaning it, and building models that can analyse it. Detailed understanding of the performance of a product or business. Providing predictions using different models and recommendations for solutions. Generating and selecting the most suitable model. Dealing with data limitations appropriately. MySQL to manage databases. Dealing with big data, using Spark, etc. Validating data.

C Programming: Using C to control/communicate with a car's ECU and sensors as part of the UWE Formula Society. Using C for embedded systems programming such as PIC microcontrollers. For signal processing. For RTOS.

Data modelling: Using Erwin Data Modeler. Using Microsoft Visio to make data visualisations.

Machine vision: Use various sensors and machine vision models to identify landmarks in the car's surroundings. Allows car to avoid obstacles effectively. Part of the 'Perception' group of the AI UWE Formula society.

Electronic Systems Design: PCB, circuit, and schematic design and interpretation (CAD/EAGLE). Soldering. Embedded systems design and programming (C/C++).

Using Test Equipment: Oscilloscopes, multi-meters, function generators, power supplies, etc.

Cloud computing/distributed computing: Cloud infrastructure, AWS, foundational Azure skills in Virtual Machines, Blob Storage, Virtual Network, Azure AD, Functions, SQL Database, etc. Multi-tenant architectures, practical uses for businesses, sustainable and efficient cloud computing, cloud computing architectures, fundamental cloud security, etc.

Computer science fundamentals: Proficient in object-oriented design, data structures, algorithm design, problem solving and complexity analysis (time and space).

Power analysis Analysing power systems. Energy efficiency and renewable energy. Building and simulating network models on PSCAD. Assessing power quality and managing it. Voltage regulation, dealing with harmonic distortion, unbalanced loads, transients, frequency fluctuations, network stability, fault detection and isolations, etc. Network planning and optimisation for current and future demands, while catering to stakeholder needs.

Embedded Systems Programming: Proficient in C, Rust, and applying these languages to control and communicate with embedded systems, including ECUs and sensors.

Operating Systems & Kernel Internals: In-depth understanding of operating systems principles, with experience in kernel internals, process management, and memory management.

Accounting: Using MS Office programmes, such as Excel and PowerPoint, to store information about a business's performance and present the information to an audience.

Software engineering: Optimising algorithms. Proficient in debugging software issues and analysing application and event logs. Adhering to the methodologies of the Software Development Lifecycle. Using methodologies such as Agile and Scrum. Creating automated test plans so that software is deployment-ready. Object-oriented programming.

HOBBIES-EXTRA INFORMATION

- **Formula racing** companies need new, creative people with a keen interest in sustainability. Competing in both the automotive AI and human-driven Formula Student competitions, where we utilised Coryton sustainable fuels, strengthened my understanding of how vital sustainability is for the future of engineering and technology.

REFERENCES

- **Dr Melvyn Smith - Machine Vision Centre co-director:** melvyn.smith@uwe.ac.uk
- **Dr Haixia Liu – Machine Learning senior lecturer:** haixia.liu@uwe.ac.uk