Alan Joe Anil

Data Science/Artificial intelligence/Deep learning/Computer vision/Machine Learning
Ready to Relocate

alanjoeanil98@gmail.com +91 7558970977 Alan Joe Anil GitHub Portfolio

PROFESSIONAL SUMMARY:

- Experience of **Machine learning** algorithms like **Linear Regression, Logistic, KNN, Support Vector Machines (SVM), Decision trees, Random Forest, Adaptive Boosting (ADABoost), Gradient Boosting, XGBoost and K-Means Clustering.**
- Strong Mathematical foundation and good in Statistics, Probability, Calculus and Linear Algebra.
- **Feature engineering** – Data quality, Missing value treatment, Data Wrangling, Label encoding, One-Hot encoding, Feature split, scaling, Overfitting and Underfitting treatment using Sampling techniques, Multicollinearity check, Cross Validation techniques, Subset selection methods, Hyperparameter tuning (grid search cv and random search cv), outlier handling, transforming variables and reshaping data using Python libraries.
- Good knowledge of Clustering algorithms like **K means, Agglomerative Hierarchical Clustering, Divisive Hierarchical Clustering, DBScan, Association Rule learning and Dimensionality Reduction like PCA, LDA**.
- Good knowledge of Deep Learning (DL) and ample hands-on with **Neural Networks, ArtificialNeural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks(RNN) and Long Short Term Memories (LSTM).**
- Basic Understanding of **Computer Vision techniques** like **Image pre-processing, ImageSegmentation, Object detection, Object recognition** etc.
- Basic Understanding of **Natural Language Processing (NLP)** techniques like tokenization, stemming, lemmatization, word2vec etc.
- Skilled in libraries like **Numpy, Pandas, Matplotlib, Seaborn, Scikit Learn, TensorFlow, Keras, OpenCV and NLTK.**
- Good at fundamentals of **SQL**.
- Basic knowledge of **Amazon Web Services (AWS) Cloud Computing** and **Google Cloud Platform (GCP)**
- Self-motivated team player with good communication and presentation skills.
- Good analytical and problem-solving skills.

TECHNICAL SKILLS:

- **Programming Languages**: Python, Machine Learning
- **Machine Learning / Deep Learning**: Supervised and Unsupervised ML, ANN, RNN, CNN using TensorFlow Keras ,Pytorch, Python (eg. scikit-learn, numpy, pandas, matplotlib)
- **Data Science & Miscellaneous Technologies:** A/B testing, Data science pipeline (cleaning, wrangling, visualization, modeling, interpretation), Statistics, Time series, Experimental design, Hypothesis testing
- **Databases**: MySQL
- **Visualization Tools and Libraries:** Matplotlib, Seaborn
- **Tools or IDE:** VS Code, PyCharm, Jupyter Notebook, Google Colab
- **Cloud platforms:** Heroku

EDUCATION QUALIFICATION:

- **Master Of Computer Application**
  Cambridge Institute Of Technology , Bangalore | 2019-2021
- **Bachelor Of Computer Application**
  Parumala Mar Gregorious College , Valanjavattom | 2016-2019

- 12th , St. Behananas HSS, Vennikulam | 2016
- 10th , St. Johns HSS, Eraviperoor | 2014

WORK EXPERIENCE:

Company	Netzwerk Academy
Role	**Data Science Intern** (December 2021 - Present)
Technologies used	Python, Machine Learning, Deep Learning, Computer Vision, NLP
Responsibilities	• Understanding the problem statement, analysing the problem, building architecture on how to solve the problem. • Collecting, cleaning, transforming and analysing the data. • Building Machine Learning and Deep Learning models. • Training and testing ML/ DL models using Tensorflow. • Learning Flask for Model Deployment.

DATA SCIENCE PROJECTS: Portfolio link

1. Mobile Price Range Prediction GitHub

Tag :	**Classification, Confusion Matrix, Hyperparameter Tuning, GridsearchCV**
Description	• Developed a binary classification ensemble model of the Decision Tree, Random Forest model to predict mobile price range in the market based on technical specs and achieved an **accuracy of 80%** on test data. • Applied Heatmap to check the correlation between features and select the important features for future implementation. • Feature engineering was carried out across screen dimensions, battery life, microprocessor strength, camera specs, and internal memory specs to see which features would perform optimally using the GridsearchCV

2. Laptop Price Predictor GitHub

Tag :	**Regression, EDA, Hyperparameter Tuning, GridsearchCV, Streamlit, Heroku**
Description	• Developed a **Regression** model of the Linear Regression, L1, L2, Decision Tree, Random Forest model to predict Laptop Price range in the market based on technical specs and achieved an **accuracy of 88%** on test data. • Feature engineering takes place and extract some columns like Memory, GPU, Screen Resolution and CPU. • Split Screen Resolution column into Touchscreen, IPS and PPI. • Got **88%** accuracy on Random forest and **Hyperparameter Tuning for Random forest** which features are encoded by using onehot encoding. • After modeling then Predict the price of laptops on the whole Dataset. • Then create a Streamlit app of this model and deployed in Heroku.

3. Cryptocurrency Price Prediction GitHub

Tag :	**Pandas, yfinance, plotly, TimeSeries, AutoTS**
Description	• Cryptocurrency price prediction by importing the necessary Python libraries and the dataset we need. • Collect the latest Bitcoin prices data from Yahoo Finance, using the **yfinance** API. • Predicting the future prices of cryptocurrency is based on the problem of Time series analysis. • The **AutoTS** library in Python is one of the best libraries for time series analysis. So here I will be using the **AutoTS** library to predict the bitcoin prices for the next 30 days.

4. Fake News Detection GitHub

Tag :	**Feature Extraction, CountVectorizer, TfidfVectorizer, Passive Aggressive Classifier, Streamlit**
Description	• Relevant Feature selection using pandas and Vectorizing them using Tfidvectorizer. • Use **Passive Aggressive Classifier algorithm** and got **93.61%** accuracy model. • Save the model and implement in the Streamlit app.

5. Traffic Sign Classification

GitHub

Tag :	**CNN**, **Pandas, tensorflow, keras, keras.layers, Flask**
Description	• Buidling a deeplearning CNN model to classify traffic signs , Train the model with 43 different traffic signs . • Using keras model and layers like Conv2D, MaxPool2D, Dense, Flatten, Dropout and mostly used activation function is **RELU.** • I used 20 epochs through which I got 96% accuracy. • Then save the model and build a **FLASK app**

6. Movie Recommendations Engine

GitHub

Tag :	**Recommender Systems**, **Feature Engineering, Missing Values Treatment**
Description	• Understanding of Missing Values Treatment. • Perform Data Analysis. • Perform Feature Engineering. • How to Manipulate the data according to the requirements. • Perform Recommendations based on the Content, Similarity, Actors, Languages and Geners

7. Omicron Sentiment Analysis

GitHub

Tag :	**Sentiment Analysis**, **Pandas, nltk, wordcloud**
Description	• The dataset that I am using for the task of Omicron sentiment analysis is downloaded from Kaggle, which was initially collected from Twitter when people were sharing their opinions about the Omicron variant . • The text column in the dataset contains the tweets done by people to share their opinions about the Omicron variant. To move further, we need to clean and prepare this column for the task of sentiment analysis. • the word cloud of the hashtags column to look at the most number of hashtags used by the people on their tweets. • 6565 sentiments opinions were **Neutral** which means that people were sharing information about the Omicron variant instead of sharing any positive or negative opinions.

PERSONAL DETAILS:

DoB : 08 October 1998

Languages Known: English, Malayalam, Tamil.