ANSHUMAAN PHUKAN

Anshuphukan031@gmail.com 7002099985
GitHub: https://github.com/Anshumaan031
LinkedIn: https://www.linkedin.com/in/anshumaan-phukan-133ba1190/

Education

Bennett University — 3rd year
Bachelor of Science: Computer science — 8.72/10

Key courses taken:

Artificial Intelligence and Machine Learning, Deep Learning, Engineering Calculus, Data Structures using C++, Probability and Statistics, Linear Algebra, Object Oriented Programming using Java, Information Management Systems, Image and Video Processing, Software Project Management, Cloud Computing, Discrete Mathematical Structures, Operating Systems, High Performance Computing

Experience

Data Analyst Intern at SAS:

(June'21 – August'21)

- Worked on enhanced analytical tasks based on SAS Viya solution in areas of Data Preparation, Visualization, Statistical analysis, and Machine Learning.
- Identified key influencing features affecting profit margin across different car types, CAR make, and model.
- Predicting Profit Margin & establish relationship between profit and key additional factors using statistical based Regression models such as Linear regression, Generalized Linear Models, Generalized Additive Model, Random Forest, Gradient Boosting, Neural Network and Selecting the champion model based on model evaluation criteria.
- Providing end results into various dashboard and reports using SAS Visual analytics
  Internship thesis:
  https://anshuphukan031.medium.com/exploratory-and-predicative-analysis-on-automobile-industry-using-sas-viya-575951b6cf2b

Computer Vision Intern at UAV Tech Pvt Ltd:

(Jan'22 –May'22)

- Working on a project which involves creating a prototype version of an autonomous unmanned drone.
- This drone will include multiple functionalities involving object detection, path navigation and precision landing.
- Implementing custom deep learning models for real time object detection for precision landing.
- Establishing complete computational graph models by combining subscriber and publisher nodes in Robot Operating System (ROS).
- Creating our custom model for object detection. The models that are to be implemented, should use lesser computational power. The final product should successfully deliver payloads from a starting point to a destination safely. The aim of this service is to supplement the existing fleet of automobiles but in a cleaner, quicker, and efficient way and hence easing the bottlenecks in the supply chain of medicine, essential and fast-moving consumer goods.

NUS Academic Internship on Data Analytics using Deep Learning: (Jun'22 –Jul'22)

Research

Brain abnormality categorization through MRI scans: The most common occurring brain disorders today stem from brain abnormalities. The general practice is to have a neurologist analyze many patient images, which becomes extremely tedious and inefficient. In this work, we propose an abnormality detection workflow revolving around advanced machine learning techniques to study the brain's medical images (X-rays/MRIs). With more than 1000 unlabeled patient images, we labeled and applied to preprocess techniques to these images (resizing, orientation, grayscale, highlighting edges through Gaussian blurring and noise removal). Then, further classifying the images into 'brain' and 'not brain' using a neural network-based classifier framework. Models like CNN, VGG16, AlexNet and ResNet were used for our research. Techniques like drop out layers and data augmentations were also included to prevent overfitting of our data. With these concrete brain images and convolution neural network architecture, classification of brain images was performed into 'Normal' and 'Abnormal.' Further steps in classification included modifications to improve accuracies and training on larger datasets to get a better fit. Currently the draft format of our research paper is ready and will look for publication next.

Paper: https://www.irjet.net/archives/V9/i5/IRJET-V9I5757.pdf

Analysis on Global Economic Freedom Index (ongoing): A detailed comparative study on a country's independence on other countries for its GDP. An index of economic freedom measures jurisdictions against each other in terms of parameters such as trade freedom, tax burden, judicial effectiveness, and more. These factors may be weighted according to their influence on economic freedom and compiled into a single score that allows for a ranking. Machine learning models can help us in predicating to what extent a nation's economy is dependent on others.

TECHNICAL SKILLS

- Machine Learning (Supervised and Unsupervised), Deep Learning, Bagging and Boosting techniques, Exploratory Data Analysis, Feature Engineering and Feature Selection, Statistical Analysis (hypothesis testing, P-value, PDF, Z-score, distribution etc.), Mathematical intuition behind ML models, Computer Vision, Robot Operating System
- **Languages**: C++, Python, SQL, HTML, CSS
- **Libraries:** Pandas, NumPy, Scikit-learn, Tensorflow, Keras, SciPy. Stats, OpenCV, NLTK, Mito (for automated exploratory data analysis in python), imblearn
- **Tools:** Tableau, VS Code, JIRA software management tool, Git GUI using GitKraken, SAS Viya (to perform entire ML Lifecycle), Jupyter notebook, Netron (for Neural Network Visualization), Google Cloud's Big Query and Data Prep by Trifacta, PowerPoint
- **OS:** Windows, Linux

PROJECTS

- End to end Real Estate Price Prediction model using python flask for deployment:

  required data was taken from Kaggle and various feature engineering techniques like Principal Component Analysis, imputations, oversampling were implemented. I used linear regression, lasso regression and decision tree algorithms to predict the price based on features like location, area, BHK etc. Used hyperparameter tuning to come up with the best accuracy score and used python flask to deploy it in a webpage. Finally, I deployed my model in the cloud with AWS.

- Air Quality Monitoring Device using Arduino:

  This project demonstrates a use case for IOT devices. MQ 135 gas sensor was used to monitor real-time presence of gases like Carbon-Monoxide (CO), Carbon-Dioxide (CO2), and other parameters like temperature, humidity. The sensor along with a potentiometer was connected to an Arduino Uno through a, which helped us display the required output in a 16*2 LCD Panel. All the equipment were connected to each other using a bread board. The current pollution measurement methodology uses expensive equipment at fixed locations or dedicated mobile equipment. Therefore, it is desirable to have access to real-time measurements to be able to quickly analyses and identify alarming levels of pollutants.

- Heart Disease Prediction model:

  Factors that influence heart disease are body cholesterol levels, smoking habits and obesity, family history of illnesses, blood pressure, and work environment. Heart disease can be predicted based on various symptoms such as age, gender, heart rate, etc. and reduces the death rate of heart patients. I tried to make a simple classification model using logistic regression to predict heart disease based on various factors. All the necessary exploratory data analysis is included in the following notebook to better understand our data.

- Carbon emission forecasting across different energy sectors using ARIMA:

  The idea behind the project is to run a prediction, analysis and forecasting system over datasets related to carbon emissions. Our focus was on certain energy sectors causing carbon emissions. It would include trends over the years, maximum and minimum contributors, etc. The analysis can be a foundation for predicting the future trends of these contributors.
  Paper: https://www.irjet.net/archives/V9/i7/IRJET-V9I7555.pdf

- Prediction of Forest Fire based on Meteorological data:

  building a predictive model using a Decision Tree classifier that can predict fire based on the meteorological data corresponding to the critical weather elements that influence the forest fire occurrence, namely temperature, relative humidity, wind speed, etc. There are 122 instances for each region and the instances are recorded for the period from June 2012 to September 2012. The dataset includes 11 attributes and 1 output attribute (class).

- Detection of Blood Cells using YOLOv5:

  YOLOv5 algorithm was used for this project to detect blood cells like Red Blood Cells, White Blood Cells, and Platelets. Object detection is the task of classifying as well as localizing the objects in an image. In Yolo, the entire image is divided into grids with each of them having its own vectors representing the coordinates of the bounding box and class type. LabelImg software was used to manually annotate the image as well as define the bounding box parameters. The images and the XML files containing the bounding box info are created as a result to train our model with custom weights. The darknet repository was cloned into the notebook to apply the object detection algorithm on our custom dataset.

Certificates

- Azure certified AI fundamentals and DP fundamentals
- Google: elements of AI certification, Google Cloud Engineering and Data Science track
- Deep Learning Fundamentals - NVIDIA
- Intel Demystify AI and OpenVINO toolkit certification
- SAS Advanced visual Analytics and Advanced Statistical Analysis
- Machine learning with Python, Introduction to Data Science - COGNITIVE CLASS, IBM
- Introduction to HTML and CSS – Michigan University, Coursera
- Introduction to Software Project Management - Simplilearn

Hobbies and other achievements

- State ranking no.1 for lawn tennis in under -12,14 category
- Represented my state and school for multiple interstates, CBSE Meets, and school National's Tennis tournaments.
- Tennis captain of my university's sports team
- Represented my state in interstate Squash Championship

Blogs and Articles

- Detailed report on Exploratory and Predictive Data Analysis using SAS Viya: https://anshuphukan031.medium.com/exploratory-and-predicative-analysis-on-automobile-industry-using-sas-viya-575951b6cf2b
- Project Concept on Online College Social Hub: https://anshuphukan031.medium.com/project-concept-on-online-college-social-hub-351a7df2f3d7
- Role of AI for Disabled People: https://anshuphukan031.medium.com/role-of-ai-for-disabled-people-b819a21eedaa
- Project report of air quality monitoring device using Arduino: https://anshuphukan031.medium.com/project-report-on-air-quality-monitoring-device-using-arduino-b2884c596372
- Evolution of Learning Automata in NLP: https://anshuphukan031.medium.com/evolution-of-learning-automata-in-natural-language-processing-nlp-ac69097f5a45
- Importance of Design and Analysis of Algorithms (DAA) for being a good programmer: https://anshuphukan031.medium.com/importance-of-design-analysis-of-algorithms-daa-for-being-a-good-programmer-ba40b4c22e16