SHANKAR PANDE
Aspiring Data Scientist
shankargpande75@gmail.com | +91-7038516378 | LinkedIn | Github
Aurangabad, Maharashtra

EDUCATION

- **Bachelor of Technology, Mechanical Engineering**
  Maharashtra Institute of Technology Aurangabad — Aug 2019 – Jul 2022 | Aurangabad, India
  Cumulative CGPA: 8.10 / 10

- **Diploma In Mechanical Engineering**
  Government Polytechnic Hingoli — Aug 2016 – Jul 2019 | Hingoli, India
  Percentage: 71.19%

TECHNICAL SKILLS

- **Programming Language** - (Python- Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn)
- **Databases** - (MySQL- Primary & Foreign Key, Joints, Clauses, Constraints)
- **Machine Learning** - (Supervised & Unsupervised Learning, Linear Regression, PCA, K-Means Clustering, KNN, Logistic Regression, SVM, Decision Tree, Random Forest, Boosting, Performance Metrics, Cross-Validation, Ridge & Lasso Regression, Bias-Variance Tradeoff, Model Optimization)
- **Deep Learning** - (ANN, CNN, RNN, PyTorch, TensorFlow, Keras, Optimizers, Activation Function)
- **Natural Language Processing** - (Bag of word, TF-IDF, Word2vec, AvgWord2vec, word Embedding)
- **Statistics** - (Hypothesis Testing, Sampling Methods, Central Tendencies, Measures of Dispersion)
- **Data Visualization** - (Power BI, Tableau, Excel, PowerPoint Presentation)

SOFT SKILLS

- Active listening, Interpretation
- Long-term thinking
- Business Sense
- Collaboration
- Team Work

WORK EXPERIENCE

Data Science Consultant, Intern - Rubixe.ai | Bangalore, India **June 2023 – March 2024**

- Collaborated with the team members to understand client requirements and business challenges.
- Constructed predictive models using machine learning algorithms, including Logistic Regression and Multiple Linear Regression, achieving an accuracy rate of 85% in attributing revenue and conversions to specific marketing touchpoints
- Worked with different technologies and different fields like Data Engineering, Big Data.

PROJECTS

1. **ITSM-Ticket-Management |** Github
   - In This Project learning models for predicting high-priority IT incidents and tickets. By evaluating previous data, the system identifies critical issues and **reducing 80% risk**
   - Managed a process re-engineering project to improve and consolidate end-to-end service processes, restructured communication flow among **10 departments** and **reduced 75% reassignment time.**
   - Used **Machine Learning** approaches to improve the **Request for Change process**. By analyzing patterns and past information, the software helps to evaluate change requests, reduce manual effort, and maintain service quality.
   - Implemented a Machine Learning technique to predict on features and achieved an accuracy rate of 90%.
   - **Technologies Used*:- :*** Python Pandas, Sklearn, Machine Learning, cleansing, matplotlib, data collection, data visualizations.

2. **Insurance Cost Prediction** | Github
   - Insurance costs have risen dramatically over the past decade in response to the rising cost of health care services and are determined by a multitude of factors.
   - The project's goal is to predict the insurance price charged to the customer using a machine learning model trained on Kaggle data.
   - To make the final forecast, addressed some values and outliers and used a variety of machine learning techniques such as **PCA, Decision Tree, SVM, KNN, Linear Regression.**
   - The impact of the project is highlighted by the achieved accuracy rate of 87% .
   - **Technologies Used*:- :*** Python Pandas, Sklearn, Machine Learning, cleansing, data collection, data visualizations.

3. **ProtugeseBank marketing project |** Github
   - The data is related with direct marketing campaigns of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit (variable y).
   - the marketing team, ways to better target customers using feature importance maps and business intuition.
   - Marketing campaign can be understood as phone calls to the clients to convince them accept to make a term deposit with their bank.
   - The best model gradient boosting classifier with optimized hyperparameters. The model's achieved accuracy rate of 79.5%
   - **Technologies Used*:- :*** Python, Pandas, Sklearn, Machine Learning, cleansing, data collection, data visualizations.

4. **FIFA – 20 Clustering |** Github
   - Business case of the Project with FIFA20 dataset need to cluster the player by their skill into certain group.
   - Unsupervised learning is a type of machine learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision.
   - This project is an unsupervised machine learning application to cluster football players based on its attribute such as defending, shooting, acceleration, marking, etc
   - KMeans did a good job in classifying players for both small and large sample size. It is relatively simple to implement.
   - **Technologies Used*:- :*** Python, Pandas, Sklearn, Unsupervised Machine Learning, cleaning, data collection, data visualizations.

CERTIFICATIONS

- Data Scientist Internship - Rubixe AI — June 2023 - March 2024
- Certified Data Scientist - DataMites — June 2023 - October 2023
- Certified Data Scientist - IABAC — February 2024
- Amazon Web Services — July - 2022