SAI KRISHNA KATHROJI

DATA SCIENTIST

(Data Science| Machine Learning |AI | Deep Learning | NLP | Text Mining | Neural Network)

Phone: 248-468-5409 Mail: [kathrojisaikrishna@gmail.com](mailto:kathrojisaikrishna@gmail.com)

**SUMMARY**

* **Data Scientist** with Data Science 8+ years of IT experience in various analytical projects, with expertise in implementing supervised and unsupervised learning techniques for data analysis.
* Proficient in **Python**, **R** programming languages, and with 5+ years of experience in **SQL**
* Implemented analysis on Academic, Pharmacy, Aviation, Bot projects by using Machine Learning models.
* In-depth knowledge of implementing statistical formulas for analysing data model predictions.
* Used NLP, text mining, TF-IDF, and neural networks for chatbots in an AI-based patient prescription bot with Deep Learning Techniques.
* Implemented models and deployed them using Flask and Heroku.
* Worked on various cloud platforms such as **BOX**, **GCP**, **AWS**, and **Azure** for data maintenance, predictions, time forecasting models, and turbine failure analysis.
* Implemented reinforcement techniques to control behaviour and encourage learning.
* Worked on ETL, exploratory data analysis, data preprocessing, feature scaling, data splitting into training, testing, and validation, and calculating scores such as R-square, F1-score, precision, recall, and model accuracy.
* Deployed models with APIs such as **FastAPI** using **HTML**, **CSS**, **JS**, and **Python** code.
* Worked with various tools and IDEs such as Oracle for SQL databases, Jupyter, Spyder, Google Colab, Python IDLE, and VS Code for Python programming, and RStudio for R programming.
* Created data visualizations using Matplotlib, ggplot2 libraries, and PowerBI, and Tableau visualization tools for presenting reports.
* Worked on **generative AI** applications using AWS Bedrock foundational models by importing boto3 packages.
* Worked on Snowflake, loading data from cloud storage to Azure databases using queries
* Worked on **Azure** Databricks, Azure Data Factory, Python, Scala, Spark, and SQL.
* Applied classification techniques such as Naïve Bayes, SVM, Decision Tree, and Random Forest with boosted techniques for decision making, and clustering for grouping models with appropriate results, including creating dendrograms and hierarchies.
* Completed end-to-end projects with stakeholders to achieve half-yearly and yearly targets.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Machine Learning Capabilities:**  Classification(Naïve Bayes, KNN, SVM, Decision Tree, Random Forest) and Reinforcement Techniques  Regression Models(Linear, Multiple, Logistic Regression)  Clustering(Hierarchical & Non-Hierarchical, K-Means Clustering)  **Data Visualization:**  Python Libraries - Matplotlib, Seaborn,  R Libraries – ggplot2, caret  Tableau, PowerBI  **Machine Learning and Deep Learning:**  Tensorflow, Keras, PyTorch, Scikit-Learn, XGboost | **Software Tools & IDEs:**  Oracle, Anaconda  Jupyter Notebook/ Jupyter Lab, Pycharm, Spyder  RStudio  **Programming Skills:**  Python Programming – Pandas, Numpy, Scipy and NLTK  R Programming – dplyr, tidyr  MySQL, PostgreSQL, SQLite, Java, Java script, Html, CSS  **Statistical Methods:**  Exploratory Data Analysis, Data Preprocessing, Predictive Analysis, Hypothesis Testing |

**CERTIFICATIONS**

Certification from EXCELR Solutions – Python and R Programming in Data Science

Certified on Azure- AI 900 | AZURE AI Fundamentals

Certified on Azure- AI 102 | AZURE AI Engineer Associates

Certified on Azure- DP 100 | AZURE Data Scientist Associate

**WORK EXPERIENCE**

**Dooney & Bourke, Norwalk, CT**  June 2024 to Present

**Data Scientist (remote)**

**Responsibilities**

* Designed and implemented machine-learning models for time series forecasting, text classification, and anomaly detection, resulting in improved predictive analytics and operational efficiency.
* Reduced model-training time by 40% through optimized data pre-processing and parallel processing techniques.
* Developed and integrated RESTful APIs for seamless data exchange between machine learning models and client applications, enhancing real-time decision-making capabilities.
* Utilized Kubernetes for container orchestration, ensuring efficient deployment and scaling of machine learning models in production environments.
* Implemented end-to-end machine learning workflows using AWS SageMaker, streamlining model training, tuning, and deployment processes.
* Worked on Gen AI for customer support using AWS for chatbots of customer queries
* Conducted real-time troubleshooting and resolved issues related to cloud infrastructure and machine learning pipelines.
* Set up and maintained CI/CD pipelines using Jenkins, enhancing deployment efficiency.
* Managed containerized applications with Kubernetes and Docker, improving resource utilization and scalability.

**Environment:** Kubernetes, Time Series Modeling, Text Classification, Anomaly Detection, Random Forest, Boosted Trees, Linear Regression, AWS (EC2, S3, DynamoDB, EMR), AWS Sagemaker

**UCB Pharmaceutical, Capgemini, Hyderabad** September 2023 to May 2024

**Data Scientist**

**Responsibilities**

* The aim of this project was to manage and optimize data integration and processing operations for UCB Pharmaceutical. Responsibilities included handling daily file management, ensuring accurate inbound and outbound processing, executing weekly and monthly tasks, maintaining database integrity, and providing status updates and communication.
* Generated monthly reports, including heat maps and analyses, based on database data. Utilized **Oracle**, PL/SQL, ET sheets, and Excel for data processing and validation.
* Worked on **Azure** Data lakes for Managing and organized structured and unstructured data
* Created and maintained daily DTR (Daily Transaction Report) after completing all tasks. Informed the PHI team post-inbound processing and proceeded with outbound processing.
* Created storage integration, configured permissions on cloud storage, loaded data from stage to table for Data warehouse using Snowflake
* Worked on Azure Databricks and implementing BI for Analysis from Azure Data Factory used Machine Learning Operations (MLOPS) for patients tracking records of pharmacies with Python
* Sent API summaries of success and failure reports via email and maintained a tracking report for all extracts.
* Worked on Deep Learning Techniques of Neural Networks on Large Data. Managed inbound and outbound processes to ensure data integrity. Utilize technical tools for data processing, validation, and reporting.
* Handled specific weekly tasks such as extra file processing on Mondays, weekend turnover reports on Tuesdays, duplicate checks on Wednesdays, NPI refresh rate checks on Thursdays, and processing all files on Fridays.

**Environment:** Oracle, PL/SQL, ET sheets, Excel, Putty, Jira, Azure Databricks, Azure Data Factory, Python, PySpark, Snowflake, Deep Learning

**GE Aviation, Capgemini, Hyderabad** March 2022 to August 2023

**Data Scientist**

**Responsibilities**

* The aim of this project is to migrate data from existing forms and sites to a cloud-based platform, while preserving the workflow processes.
* The migration will include transferring data such as forms, documents, images, and videos to **Box** cloud, ensuring that the data is secure and accessible from anywhere. The project objectives are:
* To successfully transfer all forms, sites, and data to the cloud platform.
* To preserve the existing workflow processes, ensuring that the data can be processed and acted upon in the same manner as it was prior to the migration.
* To ensure that the data is secure and accessible from anywhere, with proper backup and recovery measures in place.
* To provide training and support to end-users, so they can use the new platform effectively.

**Environment:** Cloud Libraries, Communities, Forms, Charts, Box, Sites & Workflows

**Winnee AI-Powered Chatbot, Innodatatics, Hyderabad** April 2021 to February 2022

**Data Scientist**

**Responsibilities**

* Built a **Chat bot** from scratch to Handle Customer queries, predict the category and assign tickets to the respective team.
* To provide a brief overview of the technical architecture of the model of the Chat bot which will be built to automate the query resolution system.
* Designed and developed a chatbot from scratch to handle customer queries, predict query categories, and assign tickets to the appropriate team, automating the query resolution process and improving response times.
* Implemented **machine-learning** algorithms, including KNN, Naïve Bayes, SVM, Random Forest, Logistic Regression, Gradient Boosting, NLP and Decision Tree, to accurately classify customer queries and ensure high prediction accuracy.
* After Implementing Deep Learning Techniques of Neural Networks prepared a Text Classified with clustering techniques.
* Collaborated multiple models to enhance the chatbot's performance, using model collaboration techniques to optimize the accuracy and efficiency of the query resolution system.
* Built and integrated the technical architecture for the chatbot, ensuring seamless interaction and query handling, which improved customer satisfaction and operational efficiency.

**Techniques used** - KNN, Naïve Bayes, SVM, Random Forest, Logistic Regression, Gradient Boosting, Decision Tree, Model Collaboration, NLP and Neural Network.

**Wind Turbine Failures, Innodatatics, Hyderabad** February 2020 to March 2021

**Data Scientist**

**Responsibilities**

* Analysed the performance of the wind turbine based on the location by use of historical data. Clustering the wind turbines based on the location and performance. Time series analysis for the performance of wind turbines within the cluster segment in **Pycharm** with python code with Deep Analysis.
* Analyze and predict which component/components of the turbine are going to fail. According to the predictive analysis of wind turbine failure, the maintenance (losses) cost is almost reduced by up to 60%, Data Cleaning, merging of different sensor data (in each turbine the number of sensors ranging from 150 – 250 sensors), and capturing the data from the metrology department.
* Pre-processing the data and thorough understanding of the data distributions according to the time stamp.
* Exploratory Data Analysis to understand the performance of the similar turbines based on area and trend and seasonality difference in the cluster segments, estimating the range of values of different sensors with seasonality effect.
* Time series visualizations on sensor data of major parts of wind turbines like gearbox, generator, rotor, blades. Predict the values of precision and recall by performing different Machine Learning algorithms like Random forest, XGboost, MLP with different optimization techniques.

**Techniques Used** - Time Series Visualisations, XGboost, Random Forest.

**Mobile Wallet Fraud Detection, Innodatatics, Hyderabad** June 2018 to January 2020

**Data Scientist**

**Responsibilities**

* Built a database in SQL and created a connection between **python** and **SQL**. Cleaning of data in order to remove and replace errors such as similar timestamps between consecutive transactions.
* Created derived variables such as velocity and time difference between transactions Executed feature engineering techniques to select suitable features.
* Performed stratified technique while splitting train and test data.
* Built different models using different algorithms and selected a Random Forest classifier which has higher accuracy. Deployment of model with the help of flask.

**Techniques Used** - Random Forest, Clustering and Regression Techniques, SVM.

**Academic Analysis Using Machine Learning, Innodatatics, Hyderabad** June 2016 – May 2018

**Jr. Data Scientist**

**Responsibilities**

* To Predict Overall performance of students based on the Student Attrition Rate of the university year- wise, On-Time graduation, On-Campus student employability.
* Created a Pipeline between Python and MySQL and fetch the data to **Python** from **MySQL** Database.
* Performed **EDA**, and **Data Visualization** to get insights from the data provided by the client.
* Pre-processed Data, implemented various Machine Learning Algorithms, and deployed the model using **Flask**.

**Techniques Used** - SVM, Neural Network, Naïve Bayes, And Random Forest.

**Education**

**Bachelor of Science in Statistics, Mathematics and Computer Science (2016)** – TG University, Telangana