Rohit Sahoo

Thane, Maharashtra, India | +91-9930391406 | rohitsahoo.com
rohitsahoo741@gmail.com | linkedin.com/in/rohit-sahoo | github.com/rohit-sahoo | kaggle.com/rohitsahoo

EDUCATION

Terna Engineering College, University of Mumbai **Navi-Mumbai, India**
Bachelor of Engineering in Computer Engineering (CGPI: 8.23/10) *August, 2016 – November, 2020*
Relevant Coursework: Big Data Analytics, Artificial Intelligence, Data Structures and Algorithms, Applied Mathematics, High Performance Computing, Distributed Computing, Theory of Computer Science.

RESEARCH EXPERIENCE

Alan Turing Institute **London, United Kingdom**
(United Kingdom's National Institute for Data Science and Artificial Intelligence) *September, 2021*

- Participated as a Researcher on the problem "Machine Learning Techniques to improve sleep in Dementia Patients" given by UK DRI (UK Dementia Research Institute) – Care Research and Technology Centre.
- Researched and implemented machine learning models using synthetic data generator to augment the data to enhance the sleep prediction that would help dementia patients reach their target sleep cycle.
- Received an award "*The Super-Additives*" for championing collaboration and knowledge exchange, for working respectfully and productively as a team member.

PROFESSIONAL EXPERIENCE

Tata Consultancy Services Limited (TCS) **Mumbai, India**
Assistant System Engineer (Data Scientist) *August, 2020 – Present*
Client: *Largest and leading Multinational Automotive Manufacturer*
Technologies used: Python, Machine Learning, QlikSense, QlikView, AWS, IBM DataStage

- Developed a machine learning application to identify issues in automotive parts using predictive analytics which helped in the reduction of the buyback and deployed it on Amazon Web Services.
- Implemented scalable and efficient methods for large scale data analyses and model development.
- Optimized performance and developed dashboards for applications using QlikSense that helped the client to improve their business on various Key Performance Indicators (KPIs).

Marketplace Technologies Pvt. Ltd. **Navi-Mumbai, India**
Intern *June, 2018 – July, 2018*
Technologies used: Java, Scala, Python, Apache Hadoop, Apache Spark

- Developed an application in Spark using Scala to compare the performance of Spark with Hive and SQL/Oracle.
- Optimizing of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's.

TECHNICAL SKILLS

- **Languages**: Python, C, C++, C#, Java, SQL (MySQL, PostgreSQL)
- **Data Science:** Machine Learning Algorithms, Deep Learning Algorithms, Time Series Forecasting, NLP
- **Python Libraries:** NumPy, Pandas, Matplotlib, Sci-kit Learn, TensorFlow, Keras, Spark MLlib, SpaCy
- **Big Data & Analytics Tools:** QlikView, QlikSense, IBM DataStage, Tableau, Apache Hadoop, Apache Spark
- **Cloud:** Amazon Web Services (AWS)**,** Google Cloud Platform (GCP), Microsoft Azure

INITIATIVES

Decode Engineering *February, 2021 – Present*
Founder

- Started a community to help students and professionals by providing them an easier accrual of resources along with a personal recommendation.
- Assembled a team of industry specialists to assist over hundred and fifty plus members in making the most use of the available resources.

PROJECTS

Ventilator Pressure Prediction *September, 2021 – Ongoing*

Technologies used: Python, Pandas, NumPy, Scikit-learn, Matplotlib, TensorFlow

- Implemented a deep learning model using Bidirectional LSTM to simulate a ventilator connected to a sedated patient's lung by taking lung attributes and correctly predicting the airway pressure in the respiratory circuit during the breath.
- Evaluated the model using the mean absolute error between the predicted and actual pressures, which is 0.152.

Optiver Realized Volatility Prediction *July, 2021 – September, 2021*

Technologies used: Python, Pandas, NumPy, Scikit-learn, Matplotlib, TensorFlow

- Developed an efficient model to predict short-term volatility for hundreds of stocks across different sectors using multivariate time series forecasting and ensemble learning methods such as LightGBM, XGBoost and AdaBoost.
- Evaluated the model against real market data collected in the three-month period, which yielded an RMSE of 0.28.

Capstone Project: Automated Table Extraction from PDF documents to Excel *June, 2019 – May, 2020*

Technologies used: Python, Pandas, NumPy, Scikit-learn, Matplotlib, OCR: Tesseract, Flask, React

- Developed a Machine Learning-based software that is capable of identifying tables from PDF documents and extracting the tabular information into an Excel sheet.
- Evaluated the model to determine accuracy using the F1 score, which is 0.89 for extracting information from bordered tables and 0.85 for borderless or partially bordered tables.

CERTIFICATIONS

- "TensorFlow Developer Certificate" by TensorFlow, Google (Exam Score: 25/25 test cases) *May, 2021 – May, 2023*
- "Math for Machine Learning Specialization" by Coursera and Imperial College London. (Grade: 97.75%) *Sep, 2021*
- "Machine Learning Specialization" by Coursera and University of Washington. (Grade: 97.21%) *May, 2021*

ACHIEVEMENTS

- 3x Kaggle Expert and amongst the top 5% of data scientists across the world on Kaggle platform.
- Ranked in the top 15% of world's largest coding competition "CodeVita" conducted by Tata Consultancy Services Limited (TCS) and was awarded a Full-time job offer.
- Finalist of the "PowerUp Automation" Hackathon by UiPath (global software company for Robotic Process Automation).
- Received Scholarships from Facebook AI and Bertelsmann for Udacity Deep Learning and Data Science Courses.

SELECTED PUBLICATIONS

1. **Disparity in the exploration-exploitation problem in Policy-Based Deep Reinforcement Learning for episodic and continuous environments** *Accepted (Nov, 2021)*
   Journal: International Journal of Electrical and Computer Engineering Systems (IJECES)
2. **Comparison of GANs and VAEs as methods of synthetic data generation and augmentation to enhance heart disease prediction** *Accepted (Nov, 2021)*
   Journal: International Journal of Electrical and Computer Engineering Systems (IJECES)
3. **Chapter: Machine Learning algorithms for Big Data Analytics including Deep Learning** *Accepted (Jul, 2021)*
   Book: Machine Learning Based Blockchain Technologies for IoTs and Big Data: Fundamentals, methods and applications
   Publisher: IET (The Institution of Engineering and Technology)
4. **Chapter: Building a Smart Healthcare System Using Internet of Things and Machine Learning** *Accepted (Jul, 2021)*
   Book: Nano-Robotics and Sensing – Deep Learning Applications
   Publisher: River Publishers
5. **Chapter: Image Processing** *Accepted (May, 2021)*
   Book: Computational Science and its Application
   Publisher: Apple Academic Press
6. **Paper: Auto-Table-Extract: A System to Identify and Extract Tables from Pdf to Excel** *Published (May, 2020)*
   Journal: International Journal of Scientific & Technology Research (IJSTR), ISSN: 2277-8616
   Published: IJSTR, Scopus