Jubin Mohanty

•6174683268 • +918595114827 • jubinmohanty1122@gmail.com
http://www.linkedin.com/in/jubin-mohanty-a03621a5 • https://github.com/Jubin1122 • https://jubinmohanty.com

QUALIFICATIONS PROFILE

Substantial academic training and hands-on industry experience. Graduated with a Master of Science in Data Analytics, with a concentration in Machine Learning. Aspire to bring technical skills to challenging role as **a Data Scientist, Software Developer and Data Engineer.**

- **Programming:** Works with multiple languages and scripts to design application programming interfaces (APIs) and interactive end-user dashboards.
- **Data Analysis/Data Engineering:** Skilled in analyzing data to develop business tools and improve overall systems and processes.
- **Computer Vision/Machine Learning:** I have significant research and practical experience in the field of artificial intelligence, machine learning, and the computer vision domain.

Core Technologies:

LANGUAGES:	PYTHON, R, SQL, PL/SQL, JAVA; WEB: HTML, CSS, JAVASCRIPT, JQUERY
DATABASES:	ORACLE 11G, MYSQL, MONGODB, CASSANDRA, MICROSOFT SQL SERVER, POSTGRES
REPORTING:	TABLEAU 10.5, EXCEL 2016, ORACLE ERP BB 11G
SOFTWARE & TOOLS:	ECLIPSE, SQL DEVELOPER, JUPYTER NOTEBOOK, PYCHARM, R STUDIO, VISUAL STUDIO, MYSQL WORKBENCH
SERVER MANAGEMENT & CONTAINER ORCHESTRATION TOOL:	APACHE TOMCAT, JENKINS, DOCKER, KUBERNETES, GIT
FRAMEWORKS & LIBRARIES:	NUMPY, PANDAS, SCIKIT-LEARN, MTPLOTLIB, SEABORN, GGPLOT2, ARIMA TIME SERIES, TENSORFLOW, KERAS, PYTORCH, NLTK
BIG DATA & AWS ECOSYSTEMS:	HADOOP, APACHE SPARK, AWS EMR, AWS LAMBDA, AWS EC2, AWS GLUE, AWS S3, AWS BEANSTALK, AWS REDSHIFT, AWS ATHENA, AWS CLOUDFORMATION , KAFKA, ELASTICSEARCH

EDUCATIONAL BACKGROUND

Master of Science in Data Analytics | Northeastern University, Boston, MA

Bachelor of Technology in Applied Electronics and Instrumentation | Biju Patnaik University, India, BBSR

Data Engineering Nano Degree; Machine Learning Engineering Nano Degree | Udacity

EXPERIENCE HIGHLIGHTS

Greenstand, Anchorage, AK (Part-time)
Data Scientist, Feb 2021 – Present

- Developed a computer vision model using CNN which can predict tree species and developed a model testing pipeline on AWS sagemaker.
- Currently, developing a deep learning model which can dynamically annotate images, thus will save time over manual annotation using CVAT.

Airbus, Bangalore, India
Software Developer, Oct 2020 – Present

- Contributed to the development of the data pipeline architecture for MPO tool by optimizing batch processing and utilizing pyspark modules.
- Developed and deployed a binary-classification model on AWS sagemaker for maintenance planning dossier.

Continued…

- Upgraded the legacy Airbus MPO dashboard(global) by incorporating Elasticsearch, which improved the data sync from 45 minutes to 60 seconds.

Humatics, Waltham, MA

Data Engineer, Nov 2019 – Feb 2020

- Developed optimized algorithms for an efficient ETL pipeline and designed an efficient timescale database on Amazon EC2 for ingesting more than 1 TB of data at a time.
- Devised a RNN time series model on sagemaker instance, which can predict the beacon localization and delocalization trend; thus, offers a more cost-efficient method of testing beacons performance.
- Designed infrastructure diagrams to build, deploy and automate creation of production and development infrastructure using AWS CloudFormation script.

John Hopkins Bloomberg School of Public Health/EBTC/Northeastern University, Boston, MA

Machine Learning Intern, Mar 2019 – June 2020**:** Naïve Bayes, Support-Vector Machine, Text Mining, NLP, Python

- Developed a normalization scale to categorize drug-induced liver injury (DILI) and predicted DILI categories of 503 drugs using support vector machines and naive Bayes algorithm.
- Modeled a web-crawler API to automate scrapping and mapping of literature and systematic reviews from research papers.

Daikin Air Conditioning Pvt. Ltd., Gurgaon, India

Data Engineer, Feb 2017 – June 2017

- Established an interactive and intuitive visualization dashboard for quarterly factory production using Tableau and Python frameworks such as Plotly, Django.
- Designed ETL pipelines to automate data ingestion of over 200 TB of Structured and Unstructured data.

Capgemini India Pvt. Ltd., Mumbai, India

Software Engineer, Sept 2015 – Dec 2016

- Created database objects with PL/SQL and deployed them in an Oracle ERPBB 11g production environment.
- Designed forecasting models and algorithms to drive up profits for a range of manufacturing products.
- Developed backend APIs and multiple responsive webpages with JavaScript, Jquery, HTML-5, and CSS-3 for an internal education portal.

PROJECTS

Plant Species Classification (Seedling Dataset): CNN, OpenCV, AWS Sagemaker, IceptionResNetV2, Tensorflow, MobileNetV2

I have modeled a machine learning pipeline that can perform data preprocessing, data augmentation, algorithm development, training, and model deployment into AWS. Also, the model is highly resistant to overfitting due to the fine-tunning of its hyperparameters and is highly efficient both on mobile and web applications.

Sentiment Analysis Web App: Pytorch, AWS Lambda, AWS API Gateway, AWS Sagemaker, AWS S3, NLTK, RNN, NLP, BeautifulSoup

In this application, I have designed and developed a RNN deep learning model that performs sentiment analysis on movie tweets. This model is deployed on a web application using a publicly accessible API that interacts with model endpoints. Also, in the pipeline, I have used various NLP libraries and strategies to create a dictionary of stem words from movie tweets for model training.

Plagiarism Detector: AWS Sagemaker, ANN, Gaussian Naïve Bayes, Python, AWS S3

In this project, I have built a plagiarism detector that examines a text file and performs binary classification, to label it as plagiarized or not. I covered some ongoing research topics to build the algorithms for feature extractions, such as containment and longest common subsequence using dynamic programming. Further, the application is deployed on an AWS instance, thus can be integrated in any web or mobile applications.

Facial Keypoint Detection: Computer Vision, Pytorch, Opencv, convolutional neural network, python

In this project, I have combined computer vision techniques and deep learning architectures to build a facial keypoint system, which can be used in facial tracking, and emotion recognition.

COVID19 Universal Data pipeline: Redshift, AWS GLUE, AWS S3, Apache Airflow, Python, SQL, AWS SDK

I have Conceptualized and developed a data pipeline that can parse various data formats and dynamically load data into a cloud Datawarehouse. Also, the pipeline is deployed using Apache Airflow for better automation, parallel processing, seamless data scheduling and monitoring.