Nitin Dunday Mohan

https://github.com/NITIN0601 | dmnitin97+grad@gmail.com | www.linkedin.com/in/nitindm97 | +1 9407582075

PROFESSIONAL SUMMARY

Experienced Data Engineer with 4+ years of expertise in data collection, preparation, analysis, and handling large datasets. Proficient in Python, data structures, and algorithms. Skilled in designing, implementing, and optimizing data pipelines, including ETL processes and data integration. Strong background in data engineering, data analyst, machine learning, data science and Full Stack Developer. Demonstrated track record of delivering high-quality software products and extracting valuable insights from complex datasets. Adept in agile methodologies and worked in a team environment.

EDUCATION

University of North Texas — **Denton, Texas, USA**

Master of Science in Artificial Intelligence | Machine Learning - (GPA: 3.91 / 4) — **August 2021 - May 2023**

Relevant Coursework: Feature Engineering, Empirical Analysis, Data Visualization, Fundamentals of AI, Machine Learning, Deep Learning, Software Development for AI, Bigdata and Data Science, Natural Language Processing, AI for Wearables.

Anna University — **Chennai, India**

Bachelor of Engineering | Computer Science and Engineering - (CGPA: 7.65 / 10) — **June 2014 - May 2018**

Relevant Coursework: Information Retrieval, Mobile Computing, Grid and Cloud Computing, Compiler Design, Distributed Systems, Theory of Computation, Design and Analysis of Algorithms, Software Engineering, Database Management Systems, Programming and Data Structures I & II, Internet Programming

CERTIFICATIONS & PUBLICATIONS

- **AZ-900 Microsoft Azure Cloud Fundamentals**
- **Python -** Hacker Rank
- **IBM Data Science Professional Certificate -** Coursera
- **Programming in C** – NIIT Chennai.
- **IBM Data Engineering Professional Certificate -** Coursera
- **Java Enterprise Edition Application – Level 2 -** NIIT Chennai

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, C, R-Programming
- **Database:** MySQL, MongoDB, PostgreSQL
- **Web technology:** HTML, CSS, JavaScript, typescript, ReactJS
- **Frameworks and Libraries:** Scikit-learn, NumPy, Pandas, matplotlib, seaborn, NLTK, Streamlit, PySpark, Keras, OpenAI, LLM,
- **Cloud Application:** Amazon Web Services (AWS), Azure, GCP
- **Tools:** SOAP UI, Jenkins, Postman, Wireshark, JIRA, Flask API
- **Data Visualization:** Tableau, Power-BI, Excel
- **Internal Tools:** Toro, LTDB, Dweb, Optimus, perforce
- **ETL Tools:** Azure Data Factory, Airflow
- **Big Data:** PySpark, HDFS, MapReduce, Hadoop, Hive
- **Methodologies:** Agile, waterfall,

PROFESSIONAL EXPERIENCE

Company: University of North Texas — **Denton, Texas, USA**

Role: Teaching Assistant – Computer Science and Engineering — **August 2021 – May 2023**

Responsibilities:

- Assisted professor to develop data pipeline architecture for managing student data with there class schedules.
- Conducted comprehensive data analysis on student performance based on multiple metrics such as test scores, projects, alignments, and attendance.
- Teaching undergrad students and helping the professor for preparing the teaching materials and notes for the class of 60 students.
- Led AWS Deep Racer, resulting in improved off-track race performance of students.
- Managed and maintained data storage systems such as Hadoop, Amazon S3, or SQL databases, elevating the course technological infrastructure.
- Conducted specialized seminars for students, promoting advance knowledge in cloud frameworks and data engineering pipelines in the industry.
- Multiple subjects were mentored such as Basic and Advanced programming paradigms in python, and Social Cause in Computing.
- Apart from teaching, mentoring students in their lab, project work, and clearing the doubts in the class.
- Leveraged Spark's in-memory capabilities to process and analyze extensive handle large datasets of student's information stored in S3 data lake.

Client: Norton LifeLock Software Services India Pvt. Ltd — **Chennai, India**

Role: Data Engineer – (Associate SQA Analyst) — **November 2019 - July 2021**

Responsibilities:

- Worked on various Azure services such as Azure Storage, Azure Data Factory, Azure Stream Analytics, Azure Event Hub, and Azure Cosmos DB.
- Utilized Azure Blob Storage API to store and retrieve unstructured data, such as log files and backups.
- Worked with various file formats (CSV, JSON, Avro, Parquet, ORC) in Hive to handle diverse data sources and formats.
- Experienced in manipulating/analyzing large datasets and finding patterns and insights within structured and unstructured data using pandas.
- Implemented Azure Data Factory pipelines leveraging copy activity and custom pipeline activities to ingest data from disparate sources into Azure Databricks for processing, transformation and loading into Azure Data Services.
- Utilized Azure HDInsight to perform data transformation, cleansing, aggregation, and complex analytics on batch data sets.
- Migrated the on-premises environment to Microsoft Azure cloud and ingested both structured and unstructured data into the Azure Data Lake Store.
- Created a data pipeline to move data from Azure Blob Storage to Snowflake data warehouse.
- Developed and maintained data pipelines on Azure Analytics platform using Azure Databricks.
- Utilized Azure Synapse Analytics API to perform big data analytics and reporting using SQL and Apache Spark.
- Migrated Batch ETL Crunch pipelines to in-memory Spark pipelines for improved job run times.
- Utilized Azure Stream Analytics to process real-time data streams and detect anomalies.
- Collaborated with data scientists and business analysts to understand data requirements and provide insights from the data.

- Migrated data warehouses to Snowflake data warehouse and demonstrated a high level of proficiency in Snowflake modeling and data warehousing techniques, including data cleansing, Slowly Changing Dimension phenomenon, surrogate key assignment, and change data capture.
- Created Oozie workflows to manage the execution of crunch jobs and Vertica pipelines.
- Implemented strategies to refactor, rehost applications during the migration process to optimize performance and take advantage of Azure's cloud-native capabilities.
- Created end-to-end CI/CD pipelines that automate building Docker container images, deploying containers to Kubernetes clusters, and managing rollouts from development to production environments.

Client: Symantec Software Services India Pvt. Ltd **Chennai, India**
Role: Associate Data Analyst – (Associate SQA Analyst) **July 2018 – October 2019**

Responsibilities:

- Developed Jenkins shared libraries to optimize CI/CD pipeline code reuse across multiple teams.
- Leveraged my expertise in SQL querying to extract, manipulate, and calculate information tailored to specific data and reporting needs.
- Ensured data accuracy, consistency, and completeness through thorough cleansing and preprocessing of relevant datasets
- Presented concise, actionable reports summarizing data analysis results, key findings, and recommendations to support informed decision-making
- Utilized advanced SQL techniques like joins, subqueries, and window functions to efficiently extract, manipulate, and analyze complex data from databases
- Applied statistical analysis techniques, including hypothesis testing, regression modeling, and trend identification, to derive meaningful insights
- Collaborated cross-functionally, leveraging analytical expertise and insights, to drive data-driven initiatives and support decision-making processes
- Created visually engaging data visualizations, such as charts, graphs, and dashboards using tools like PowerBI and Excel, to effectively communicate data insights
- Leveraged SAS to efficiently extract, manipulate, and analyze large, complex datasets from disparate sources, significantly enhancing data preparation and analysis through the creation of automated SAS macros, which reduced repetitive tasks, minimized errors, and boosted efficiency.
- Developed procedures, functions, and packages to create comprehensive summary tables tailored to specific data analysis and reporting needs
- Collaborated closely with data analysts and data science experts to define validation rules and establish automated processes ensuring data quality and integrity
- Meticulously documented data sources, methodologies, and assumptions, ensuring transparency, reproducibility, and compliance with data analysis practices.
- Automated builds deployment in production server using python, reduced the manual posting of builds by 80%.

PROJECTS

MLOPS Azure-E2E **January 2023 – May 2023**

- Built an end-to-end pipeline on Azure for deploying machine learning models
- Automated model retraining, deployment workflows using Azure ML and Airflow.
- Utilized the IBM Telco Customer Churn dataset to develop a model that effectively reduced the number of customer service representatives (CSR) based on specific requirements.
- Executed various pipelines including data preprocessing, model training, model deployment, MLOps workflow, batch score updates, and real-time UI display of model process status

Technical Skills: *Azure, Machine Learning Model , Feature Training, Pipeline, Github, TensorFlow, Data Bricks, Azure API, DBX Tool,*

Gesture Motion **September 2022 – December 2022**

- Developed a hand gesture recognition system with 95% accuracy, enabling seamless file navigation and non-verbal communication
- Implemented a machine learning model and computer vision techniques for robust gesture tracking and detection with 97% efficiency
- Integrated text-to-speech functionality with 99% conversion accuracy, providing audible voice output for selected text
- Significantly improved accessibility and user experience for individuals with disabilities or unique communication needs, enhancing their communication efficiency by 80% and overall system effectiveness by 85%

Technical Skills: *Html, JavaScript, CSS, OpenCV, TensorFlow, PoseNet, HOG(Histogram Oriented Gradients.,*

Question and Answering using NLP (BERT, AlBERT,GPT2,Roberta) **January 2022 – May 2022**

- Pre-trained machine learning models were used to predict the answer from the given document or link provided by the user.
- The first and last 100 words were visualized using python seaborn, and word cloud to identify the frequency of repeated words used in the document.
- Compared the model with multiple machine learning techniques and obtained an accuracy of 80%.
- The project was deployed on the web server using HEROKU.

Technical Skills: *Machine Learning Models: (BERT, AL-Bert, GPT-2, Roberta), Python, Seaborn, word cloud, Confusion matrix, Heroku.*

Crypto Currency using Data Visualization **September 2021 – November 2021**

- Created interactive dashboard showing insights on cryptocurrency markets using Python, JavaScript, Plotly, and CoinBase API
- Visualization dashboard was created to graphically represent the market supply based on the targets using 80 different currencies based on statistics.
- Coin Base API was used to fetch the current, historical market data on a particular cryptocurrency.
- Each chart had 5 to 6 different types of filters to enhance the view of market flow for a particular stock based on the timeline.

Technical Skills: *CSS, D3, JavaScript, Tableau, python, Bar chart, pie chart, Scatter plot, Bubble chart using seaborn, matplotlib, ploty, Boken.*

AWARDS, ORGANIZATIONS AND VOLUNTEERING

- Active member of the India Student Association (ISA) at the University of North Texas, serving as Treasurer since December 2021.
- Member of American Indian in Science and Engineering Society (AISES), a national, nonprofit organization dedicated to the academic excellence of different groups in science, technology, engineering, and math (STEM) studies and careers.
- Active Membership in CSI (Computer Society of India).