HITEN REDDY KAMBAM

(316)-226-2004   hitenhanni@gmail.com | hxkambam@shockers.wichita.edu |
https://www.linkedin.com/in/hiten-reddy-kambam-06ab1820a

SUMMARY

Data Science enthusiast with 5 years of experience as Software developer and Data Engineer in Analytics technology. Skilled in Extract, Transform and Load (ETL) process (analyzing, organizing, and processing) on huge data sets and extracting information and presenting it in various visualization forms. Decent experience in Python, Azure, Object Oriented Programming, Natural Language Processing, Machine Learning and Deep Learning techniques.

EDUCATION

Master of Science in Computer Science
Wichita State University, Wichita, KS
GPA: 4.0/4
May 2024

Bachelor of Technology in Computer Science Engineering
Jawaharlal Nehru Technological University, Anantapur , India
GPA: 7.5/10
May 2019

TECHNICAL SKILLS

Programming Languages: C#, Python (Pandas, NumPy, SciPy, scikit-learn, Matplotlib, Seaborn), Pyspark, R, D3, Java Script.
Web Technologies: Angular, Php, HTML and CSS.
Frameworks: Spring MVC, Spring Boot, Hadoop, Micro-Services, Spark, Kafka, Map-Reduce, Druid, Airflow.
Databases & Data Warehouses: RDBMS, SQL, MySQL, Oracle, PostgreSQL, Enterprise Data Warehouse (EDW), Parallel Data Warehouse (PDW).
Version Control & Process Tools: Git, Team Foundation Server (TFS), Putty, Confluence, JIRA.
Visualization Tools: PowerBI, Azure, Databricks, SSMS, Quicksight, D3, DAX query, Tableau, Python (Seaborn, Matplotlib).
Analytical Techniques/Tools: Access, Clustering, Regression (Linear, Logistic), Random Forest, Data-Preprocessing, Data Pipeline, Word embeddings (TF-IDF, word2Vec), SAP Business Objects (SAP BO), Data Build Tool (DBT), Informatica.
Statistical Skills: Exploratory Data Analysis, Neural Networks, Dimension reduction, Hypothesis Testing.
AWS Technical Stack: AWS Lambda, AWS SQS, SNS, EC2, Redshift, Athena, S3, Glue, AWS Step Functions, Dynamo DB.

TECHNICAL EXPERTISE

- 5 years of experience in scripting languages like SQL, PostgreSQL, Python to frame effective queries.
- Experienced in data curation, data modeling on SQL and PostgreSQL databases, database management, and data validation using great expectations python library.
- Experienced in working on Microsoft PowerBI and Tableau which includes loading data from various data sources and then performing analysis on it and then finally creating useful Dashboards and reports.
- Experienced in AWS and Google Cloud Platforms (GCP), BigQuery.
- Designed and implemented database designs (data modeling on SQL with different schemas) for client requirements.
- Experienced in working with agile team and methodologies.

PROFESSIONAL EXPERIENCE

Graduate Research Assistant   Aug 2022–Present
Wichita State University | Wichita, Kansas

- Worked on the Migrating the American Community Survey census data from Oracle to PostgreSQL database systems.

- Worked on University Transfer Office (UTO) project where in a tableau dashboard is created to track the progress of each Wichita State University campus in the student transfer process.
- Worked on AML project to analyze resilience among cancer patients. Reports are created using RedCap software.

Graduate Teaching Assistant   Jan 2024–Present
Wichita State University | Wichita, Kansas

- Engaged in collaborative research projects that required sophisticated data analysis and algorithm development in Python, contributing to the publication of findings in peer-reviewed journals and presentations at international conferences.
- Championed a research endeavor aimed at elucidating privacy policy vulnerabilities inherent in Android applications, leveraging the Appium tool in conjunction with the Python programming language.
- Interacted with postgraduate students, offering direction in Software Analysis Methodologies and Advanced software Engineering.

ETL Data Engineer (PROJECT SHELL)   June 2020–Aug 2022
WIPRO | Banglore , India

- Designed, developedand maintained an automated and optimized ETL pipeline for continuous monitoring of incoming raw Invoice data quality and to perform transformations on raw data.
- Designed and developed Microsoft PowerBI dataflows, datasets, and dashboards to give the business side insights about customers' contract lengths, spend behavior and analytics of clients, discounts ShipSigma offered to clients.
- Assisted in production support by resolving source data issues and refining transformation rules to align with center objectives.
- Designed data schema and operated internal data warehouses and SQL/NoSQL database systems.
- Involved in development, testing, and integration of ETL routines (including those for geospatial data) using ETL tools and external programming/scripting languages as necessary.
- Established ETL workflows to support projects across the center; identified and evaluated data sources for quality, reliability, and appropriateness; and migrated and integrated datasets as needed for analysis.
  Assisted in the design of data collection and analysis efforts with a focus on preventing redundancies, improving quality, and developing systems to support longitudinal data collection and analysis.
- Helped the team with troubleshooting, researching the root cause, and thoroughly resolving defects in the event of a problem.
- Designed, developed, and maintained several Dashboards and reports on PowerBI to present the meaningful information to the stake holders as per their requirements.
- Involved in designing and developing ETL processes using SQL Server Integration Services (SSIS) and Streamsets which has various data sources as inputs.
- Performed data transformations and analysis using SQL Server Analysis Services (SSAS) Tabular cube models.
- Created 'InfyME' web search application using Spring Boot and Micro services.
- Performed data extraction and insertion into MySQL using Python pandas, transformation in Data Build Tool (DBT), DataValidation using Great Expectations, and visualization via Microsoft PowerBI.

Software Engineer (PROJECT MAHINDRA AND MAHINDRA)   May 2019–June 2020
WIPRO | Banglore, India

- I streamlined the Redshift table subscription process, which had previously been performed manually by the team. To accomplish this, I used CICD pipelines to deploy python code into the AWS console. This involved leveraging AWS Lambda, SQS, and SNS services, resulting in a 20-hour reduction in team workload per week.
- To ensure the dataset was fit for purpose, I utilized Amazon Glue crawler and Athena services to perform data parsing, refinement, analysis, and transformation. Once the dataset had been processed, I provided the downstream Machine Learning team with the necessary data features to enable them to build models capable of predicting which products were dangerous.
- Conducted an analysis on unstructured data pertaining to various products listed on Amazon's website. This analysis involved identifying error records, testing the quality and the correctness of data records.
- Developed Natural Language Processing techniques to process customer complaints data that was in about 10 different languages, translating them to English using both IBM Watson and Azure cognitive services. Subsequently, I performed an analysis on the translated complaints and used the information to develop a PowerBI dashboard and report that
- With the help of IBM Watson, I employed classification techniques to categorize the customer feedback data into five different categories: product praise, marketing feedback, displeasure with product labeling, displeasure with product packaging, and displeasure with product design.

PROJECTS

- **Political Party Prediction:**

  Built a feed forward deep neural network to predict the political party from the twitter tweet text. Tried multiple models with different hyper parameters and various regularization and optimization techniques and then tabulated the results (accuracies) of different models.

- **Face Identification Using Convolution Neural Networks (CNN):**

  Built a Convolution Neural Network (CNN) using Eigen faces to identify the faces. Used Principal Component Analysis (PCA) technique to flatten the face images to vectors.

- **Recreated Dr. Jon snow's map of cholera Outbreak:** Recreated Dr. John Snow's map of Cholera

  outbreak in 1854 in an interactive visualization version using D3 framework and Java script programming language.

- **Employee Attrition Prediction Model:**

  Built Employee attrition prediction model using machine learning algorithms to predict attrition rate by various factors like salary, Job satisfaction, work hours.
  Performed turnover analysis by using Python's Scikit-Learn library. Used Logistic Regression, Random Forest, and Support Vector Machine as classifier for employee attrition and measured the accuracy of models that are built.

- **Quantify a car:**

  CreatedPowerBI interactive dashboard to understand variation of price with more than 10 different features of car. Built a linear regression model using Pyspark to predict car price and deployed it on Heroku cloud platform.

- **Next Word Prediction:**

  Developed a next word prediction model by using RNN such as LSTM, biLSTM and transformer architecture to suggest next word of a review for a user based on previous customer reviews.

  ## CERTIFICATIONS

- Python for Data Science, AI & Development by IBM

- Data Visualization and Dashboards with Excel and Cognos by IBM

- Introduction to Data Analytics by IBM

- AWS certified cloud practitioner by AWS