PURNA SAI SOWMYA KARPURAPU

(317) - 772–2942 |purnasaisowmya@gmail.com|pukarpu@iu.edu|
https://www.linkedin.com/in/sowmya-karpurapu/

SUMMARY

Data Science enthusiast with 4.5 years of experience as Business Intelligence 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, C#, Object Oriented Programming, Machine Learning and Deep Learning techniques.

EDUCATION

- **Master of Science in Applied Data Science**
  Indiana University-Purdue University, Indianapolis (IUPUI), Indianapolis, IN

  GPA: 3.92/4
  December 2022

- **Bachelor of Technology in Electronics & Communication Engineering**
  Jawaharlal Nehru Technological University, Kakinada, India

  GPA: 8.84/10
  May 2018

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, Quicksight, D3, Power query, DAX query, Tableau, Superset, 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

- 4.5 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

Data Engineer Intern
Amazon.com | Austin, Texas

May 2022 – Aug 2022

- Designed data schema and operate internal data warehouses and SQL/NoSQL database systems

- Designed, implemented, and automated deployment of the distributed system for collecting and processing log events from multiple sources

- Provided assistance to the team with troubleshooting, researching the root cause, and thoroughly resolving defects in the event of a problem

- Automated the Redshift table subscription process which saved the team 16 man work hours per week.

Data Engineer Intern   Jan 2022 – May 2022

ShipSights | Indianapolis, Indiana

- Designed, developed and 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.
- Performed data extraction and insertion into MySQL using Python pandas, transformation in Data Build Tool (DBT), Data Validation using Great Expectations, and visualization via Microsoft PowerBI.

Graduate Assistant   Jan 2021 – Dec 2022

School of Informatics and Computing, IUPUI

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

Associate Analyst (Microsoft Business Intelligence Developer)   Jun 2018 – Dec 2020

Infosys Limited | Mysore, India

- Designed, developed, and maintained several Business Intelligence Solutions and then built 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.

PROJECTS

- **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.

- **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: (*Checkout the Project*)**
  Recreated Dr. John Snow's map of Cholera outbreak in 1854 in an interactive visualization version using D3 framework and Java script programming language.

- **Quantify a car: (*Project Documentation*, *PowerBI Dashboard*)**
  Created PowerBI 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.