Lingampally Priyanka

+1408-916-8532 | Priyankalingampally@my.unt.edu | www.linkedin.com/in/lingampally-priyanka-1858ba105

EDUCATION

- **Master of Science, Data Science** (*GPA: 3.8/4.0*)   *University of North Texas*   Aug 2019 – Present

TECHNICAL SKILLS

Technologies/Tools:	Python (NumPy, Pandas, Scikit-learn, Sci-Py), Excel (advanced), Java, HTML, CSS, JavaScript, XML, Xilinx ISE9.2i, Bootstrap
ML Algorithms:	Regression, Classification, Clustering, SVM, Random Forest, Neural Networks, Cross- Validation, Ensemble methods, Principal Component Analysis
Big-Data Technologies:	Hive/Impala, Apache Spark, Hadoop, Sqoop, HBase, Pig, Kafka, MapReduce, MongoDB
AWS services:	EC2, S3, VPC, CloudFront, AWS Lambda, CloudWatch, CloudFormation Templates
Database:	Oracle, MySQL, DB2
Data Visualization:	Tableau, Power BI, Matplotlib, Excel-Pivot Charts
Operating Systems:	Microsoft Windows, LINUX, MAC OS X.
Data Mining Tools:	SAS Enterprise Miner, Rapid Miner.

RELATED COURSEWORK

Course work: Tools &Techniques of Big Data Analytics, Data Mining, Machine Learning, Data Visualization, Data Modeling, Data Analysis, and knowledge discovery.

EXPERIENCE

Data Engineer   **Accenture Solutions Pvt. Ltd., Hyderabad, India**   **July 2017 – June 2019 (2 years)**

- Experienced in data preprocessing, data analysis, machine learning to get insights into structured and unstructured data.
- Analyzed data using Python, SQL, Microsoft Excel, Hive, PySpark, Spark SQL for Data Mining, Data Cleansing, Data Munging and Machine Learning.
- Analyzed SQL scripts and designed the solutions to implement using PySpark.
- Used JSON and XML SerDe's for serialization and de-serialization to load JSON and XML data into Hive tables.
- Used SparkSQL to load JSON data and create Schema RDD and loaded it into Hive Tables and handled structured and unstructed data using SparkSQL.

ACADEMIC PROJECTS

BIG DATA ANALYTICS USING APACHE SPARK ENVIRONMENT:
- Analysis of Movies lens data through Spark RDDS, Spark Data Frames, Spark SQL, broadcast variables, accumulators and finding insights such as highest rated movie, most popular movie, most popular Superhero, and finding similarity between movies.
- Developed and ran Spark jobs quickly using Python.
- Scaled up to larger movie data sets (10 million ratings) using Amazon's **Elastic MapReduce** service.
- Translated complex analysis problems into iterative or multi-stage Spark scripts.

TEXT CLASSIFICATION WITH NAÏVE BAYES:(Python & Machine Learning)
- Implemented Naïve Bayes classification from scratch for sentiment analysis of customer reviews and achieved accuracy of 94%. Role: Worked on estimation of probabilities, log probabilities, Frequency counting, estimating parameters for Naïve Bayes classifier, cross validation, error analysis using Python.

DATA MODELING-SQL DATABASE STRUCTURE:
- Created a SQL database structure for Scott Brown Properties of Denton using SQL join queries to get a better view of the full property management in the county and developed database architecture.

DATA VISUALIZATION -ANALYSIS OF GOOGLE PLAY STORE: (Python, PowerBI):
- Performed Data Cleaning, Data visualization (PowerBI) on Google Play Store Data where I analyzed the data to help the developers understand what kinds of apps are likely to attract more users. Also used python libraries such as pandas, seaborn, matplotlib for certain visualizations. Created pivot tables and pivot charts using Microsoft Excel.

DATAMINING - PREDICTING MENTAL HEALTH IN TECH SURVEY: (SAS ENTERPRISE MINER AND PYTHON)
- Experience in working with SAS enterprise miner 15.1 software. Analyzed the 2014 U.S Hospital mental health data performed data mining techniques like Data cleaning, Decision Tree ,Regression, Clustering ,Neural Networks and Association rules to get insights like if the employer provides necessary resources for their employees to seek help for their mental health issues .visualized key factors using Power BI software.