Sanni Kumar Gupta

Block-E, Flat. No 516, Yadav Apartments

Near Ram Medicos, DLF Phase 3, Gurgaon, Haryana

Pin No: 122001

Contact No: +918142702000

Email Id: sannigupta10792@gmail.com

Profile Summary

- Experience in working with AWS services like: Amazon S3, Amazon EC2, Amazon EMR, Athena, Glue, Amazon System Manager, EMR Notebook.
- Good understanding of Algorithms, Data-Structures Object-Oriented Programming (Oops).
- Experienced in writing SQL Queries, functions, packages, tables, views, triggers and Strong experience in Data warehouse concepts.
- Implemented discretization and binning, data wrangling: cleaning, transforming, merging and reshaping data frames
- Experience with modern big data tools such as PySpark, Hive & Airflow.
- Skilled in optimize, debugging/troubleshooting issues in complex applications.
- Experience in working with different operating systems WINDOWS and LINUX.
- Experience in continuous build and version control systems like Jira, CICD pipeline.
- Excellent Interpersonal and communication skills, efficient time management and organization skills, ability to handle multiple tasks and work well in a team environment.
- Basic knowledge of Django Framework and REST API.
- Knowledge of Machine Learning Algorithms like Linear Regression, Logistic Regression, Decision Tree, Embedded Tree, Random Forest etc.

Technical Skills

- **Languages:** Python.
- **Operating Systems:** Windows, UNIX.
- **Databases:** MySQL, SQLite
- **Python Libraries:** NumPy, SciPy, Pickle, Pandas,SciKit Joblib, Boto3 etc.
- **Big Data Tools:** PySpark & Hive.
- **Web Servers:** Amazon S3, Amazon EC2, Amazon EMR, Amazon System Manager, EMR Notebook, Apache Airflow, PySpark & Hive.
- **Development Tools:** Spyder, PyCharm, Sublime Text.

Professional Experience

- Senior System Associate (May 2015 - July 2018): IBM India Pvt Ltd, Hyderabad.
- Senior Software Developer (Nov 2018 - Dec 2019) : Future Focus Infotech Pvt Ltd, Delhi
- Application Development Senior Analyst (Dec 2019 -Till Date) : Accenture India Pvt Ltd. Sector 21, Gurgaon, Haryana

Projects

Project 1

Client: RBS (Royal Bank of Scotland)

Division: Banking-RMA (Risk Modelling Analysis)

Project: Wholesale IFRS9(International Financial Reports Standard)

IFRS 9 classification and measurement:
Classification determines how financial assets and financial liabilities are accounted for in financial statements and, how they are measured on an ongoing basis. IFRS 9 introduces a logical approach for the classification of financial assets, which is driven by cash flow characteristics and the business model in which an asset is held. This single, principle-based approach replaces existing rule-based requirements that are overly complex and difficult to apply. The new model also results in a single impairment model being applied to all financial instruments, thereby removing a source of complexity associated with previous accounting requirements.

Pre-Processing - (Steps have been coded in PySpark)

- Extraction of data from input sources
- Read Reference Data
- Join Input Data from different input sources
- Join Input Data with Reference data
- Update Column Names – Model Name and Version
- Rename column names
- Create New Column/Flags

Model Calculation - (Included steps with complex Calculation and Statistical Functions, these steps coded in Python.)

- Z Batch Calculation
- PD (Probability of Defaults) Calculation - The probability of default (PD) is the probability of a borrower or debtor defaulting on loan repayments.

- Delta DD Calculation
- DD PIT Calculation
- PD PIT Calculation
- An Value calculation
- BRPT, MPRT and CRPT Matrix Calculation
- Term Structure Calculation
- PD Cumulative, Forward Marginal and Scalar Calculation

- LGD (Loss Given Defaults) Calculation - Loss given default (LGD) is the amount of money a bank or other financial institution loses when a borrower defaults on a loan, depicted as a percentage of total exposure at the time of default
  - Quantile Calculation
  - LGD PIT Calculation
  - LGD Term Structure Calculation
  - LGD Scalar Calculation
  - Counterparty Calculation

IFRS9-Schedule Task –Using AirFlow

- Create DAGs for all calculator (Controller dag (Master dag) and Trigger dags)
- Airflow Operators like Bash, Pythonbranch, TriggerDagRun operators which help us to run our script over EMR cluster.
- XCom Push/Pull and Variables.
- Pyspark and HIVE to create tables and trigger dags job on working day of month.

Project 2

Client: Diabetic Association of India

Division: HealthCare

Project: Predict the diabetes status of a female patient given their health measurements by Pima Indian Diabetes dataset.

- We'll load up the data set into a Pandas Data Frame.
- Analysed data features about categorical and numerical features. Identify input and output features and data category.

- Performed data cleaning, transform categorical data into numerical handling missing data by using Scikit-learn, NumPy and Pandas packages.
- My approach for detecting outliers KDE (Kernel Density Estimation) by using Scikit-learn, stats models, NumPy and SciPy packages.
- On basis of predicted data, we have checked the model accuracy by using different methods (Classification Accuracy, Null Accuracy, Classification Error, Sensitivity and other analysis on Confusion Matrix) by using Sklearn and NumPy packages.
- Adjust the classification threshold.
- ROC curves and Area under the curves (AUC).

Education

- University: Uttar Pradesh Technical University
- College: GLA Institute of Technology and Management, Mathura (U.P)
- Degree: B.Tech. (Electronics and Communication), 2010-2014 (Full Time)

Certifications

- Robotic Process Automation – Uipath
- Python Certification – HackerRank