Shashidhar Reddy Daida

[www.linkedin.com/in/shashidhardaida](http://www.linkedin.com/in/shashidhardaida)| [shashidharreddydaida@gmail.com](mailto:shashidharreddydaida@gmail.com) | 2147358060

**Education**

**The University of North Texas**, TX, USA Expected Aug 2021

*Master’s in Data Science* Cumulative GPA: 4/4.00

**CVSR School of Engineering,** Hyderabad, IndiaSep 2012 – May 2016

Bachelor’s in Electrical and Electronics Engineering GPA: 3.6/4.00

# Work Experience

**The University of North Texas**, TX, US —*Research Assistant* Jan 2020 – Till date

* Actively participating in a Research Project called **Characteristics of Information Spreading across nations**
* Developing an efficient and reliable **Data pipeline** to pull the data from Twitter and process the data for further data analysis and predictive modelling.
* Conducting several **Data Engineering** experiments to understand and improve the metrics of the user content using **Python**
* Implementing **ML** classification by performing sentimental analysis using **Hive** to predict the most trending hashtags by the day, Finding whether the given tweet has positive sentiment or negative sentiment and trying to improve the model precision to 80%

**Diebold Nixdorf India Pvt Limited,** Mumbai, India — *Big Data Analyst* May 2018 – Oct 2019

**Project:** Profession Services - Retail

* Processed and analyzed the data using **Spark, Python, Kafka and Hive**. Automated **ETL** processes using shell script and created hive bucketing, partitioning tables
* Involved in frequent improvement of data warehouse by troubleshooting the daily ETL processes and deep dive into the user generated data from Retail stores in order to derive the sophisticated data models
* Strategized the campaigns with the market basket analytics with the automation of user purchase patterns and live event data
* Improved reliability and efficiency of supply chain database system to 86% through analysis of vulnerabilities using **TPAnalyze**, reduced downtime and increased order processing frequency

**Cognizant Technology Solutions,** Mumbai, India — *Data Engineer* Dec 2016– May 2018

**Client:** Sanofi Pasteur

* Designed the workflows using **Informatica Unix commands and Teradata utilities**
* Worked on various ETL tools like **Tableau** and Power BI to build the data pipelines, which are used to collect the raw data of drugs, process that data and transform to be ready for analysis
* Performed POC’s like creation of various workflows, Transformations from migrating the project from informatica power center to **Google Cloud Platform** replacing Informatica with On-premises Post-Gre SQL
* Designed and built 150+ workflows using Python scripts, Stored procedures, macros, user defined functions and views.

**Academic Projects:**

**Predict the Accident Severity in USA (Machine Learning, Python, Apache spark)**

Developed a Machine Learning classifier to classify the accident severity on scale of 1 to 5 using Apache spark MLib. The data is downloaded from kaggle

**Technical Skills:**

**Programming Languages** Python, R-programming, SQL

**Big Data** Hadoop, Pig, Hive, Kafka, Spark, HDFS, MapReduce, Flume

**Cloud Technologies** AWS (S3 storage, Ec2 service)

**Analytical Tools** Tableau, PowerBI, Ms Excel, Pivot tables, SAS Enterprise Miner

**Databases** MsSQL, MySQL, MongoDB, Cassandra, Snowflake, AWS Redshift, SAP HANA

**Machine Learning** Linear Regression, Logistic Regression, Decision Tree, Random Forest, SVM, Classification, Clustering,

KNN, PCA, Neural networks

**Honors & Activities:**

* Received certificate of appreciation for improving reliability of supply chain database system at Diebold Nixdorf
* Received a scholarship of $6000 during graduation for exceptional performance as Research Assistant
* Worked as coordinator and fund raiser for annual technical fest at college during bachelor’s degree