

Ashish Sharma

214.986.0577 | axs170025@utdallas.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

The University of Texas at Dallas

M.S., Business Analytics (Data Science)

August 2018 - May 2020

GPA 4.0

Jaypee University of Information Technology, Solan, India

Bachelor of Technology, Electronics and Communication Engineering

July 2010 - June 2014

GPA 3.88

SKILLS AND COMPETENCIES

Technical Skills: Python, R, SQL, Machine Learning, Scikit-learn, Natural Language Processing (NLP), TensorFlow, keras, Deep Learning, Tableau, Power BI, SQL Server, Microsoft Azure, MLLib, SAS, Hadoop, Hive, Apache Spark, PySpark, Big Data, Visual Studio Code, node.js, JavaScript, Statistical Modeling, Decision Tree, Regression, Random Forest, Gradient Boosting, Advanced Microsoft Excel, Microsoft Office Suite, Pandas, NumPy, Seaborn, Matplotlib

BUSINESS EXPERIENCE

Data Science Intern – Divergence.ai, Dallas, TX

June 2019 – Present

- Built a POC for a chatbot using Microsoft bot framework and SQL server to bridge the communication gap between the customers and the company, by making the intelligence reach out to every customer
- Designed the bot using node.js and JavaScript, and integrated it with the customer database, and LUIS an Azure cognitive service that provides a Natural Language Processing system for the intent and entity recognition
- Integrating Power BI embedded with the chatbot, so that it returns visualizations based on the questions asked by the customer
- Enabling row-level security to the visualizations and using NLP to prioritize the questions that the bot is not able to answer yet

Data Analyst - Cognizant Technology Solutions, India

January 2015 – February 2018

- Extracted historical auto and home insurance data of 1.2 million customers using SQL to analyze the likelihood of a prospective insured that have received a quote purchasing insurance from the company
- Prepared data for analysis by cleaning and transforming the extracted data, and performed descriptive analysis using python
- Built a classification model to predict the purchase likelihood of a customer, who is in between quote and policy purchase stage and identified the variables that impacted customer conversion rate significantly
- Implementation of the proposed business recommendations based on the model helped the client in improving the customer conversion rate by 18%

HACKATHONS AND ACADEMIC PROJECTS

Informa Analytics Challenge: LA Restaurant Data Analysis (*Machine Learning, Python, Tableau, Clustering*) [GitHub](#) - Winner

- Built a Naïve Bayes classifier to predict the health grade of the restaurant in the 88 cities of the Log Angeles county using only its name, address, and zip code with an accuracy of 63%
- Clustered the data using K-means clustering and visualized it on Tableau to understand the effect of seasonality, restaurant type, violation code, and demographics on the grade of the restaurant

UNT Hackathon: Sign Language Interpreter (*TensorFlow, Keras, Python, OpenCV*) [YouTube](#) - Winner

- Implemented a sign language interpreter to help more than 70 million people across the world, having a condition of hearing impairment, with their daily communication needs using Deep Learning
- Created a dataset of over 102k greyscale images of 44 gestures in American Sign Language using OpenCV library in Python
- Trained a convolutional neural network using Keras on the images utilizing TensorFlow backend, added 6 different convolutional and pooling layers, to predict ASL hand gestures in live video feed with an accuracy of 96.2%

Minimizing Churn Rate Through Analysis of Financial Habits (*Pandas, Machine Learning, Scikit-Learn, SAS*) [GitHub](#)

- Processed and cleaned the product-related customer data of mobile application used for finance tracking, to perform Exploratory Data Analysis and built a machine learning model using Logistic Regression to predict which customer may churn
- Validated the results using K-fold Cross Validation and performed feature selection to get the most important features, rebuilt the model with the best parameters achieving an accuracy of 62%

Data Analysis of PUBG game (*Spark, Hive, Impala, MLLib*) [GitHub](#)

- Created hive tables for PUBG game data and performed descriptive analysis by running ad-hoc queries in Impala to determine the effect of factors such as killings, distance traveled, weapons acquired on chances of winning the game
- Built models In Spark using PySpark to predict the winner of the game, achieved >91% accuracy by using linear regression and random forest, tuning hyperparameters with grid search and cross-validation