# Ashish Sharma

214.986.0577 | axs170025@utdallas.edu | www.sharma-ashish.com | LinkedIn | GitHub

## EDUCATION

**The University of Texas at Dallas** **August 2018 - May 2020**
*M.S., Business Analytics (Data Science)* **GPA 4.0**
**Jaypee University of Information Technology, Solan, India** **July 2010 - June 2014**
*Bachelor of Technology, Electronics and Communication Engineering* **GPA 3.88**

## SKILLS AND COMPETENCIES

Python, R, SQL, Machine Learning, Deep Learning, CNN, GitHub, Scikit-learn, Pandas, Microsoft Azure, Azure Databricks, Decision Tree, Random Forest, Neural Networks, TensorFlow, keras, Natural Language Processing (NLP), Gensim, Tableau, Power BI, MS SQL Server, SAS, Hadoop, Hive, PySpark, JavaScript, Microsoft Excel

## BUSINESS EXPERIENCE

***Data Scientist Co-op*** – Cotiviti, Atlanta, GA **August 2019 – Present**

- Worked with health plan client data set with more than 100 million claim lines and 2000+ features to help solve the problem of provider abrasion caused by high levels of denied payments, using the analytic approach
- Built AI/ML models on Azure ML using XG boost and Random Forest algorithms, achieving >80% precision
- Presented actionable insights to the business using predictive model output, showing a roadmap of where and how to tweak the business rules to reduce the appeal and overturn rate, suggesting quarterly savings of $3MM
- Developed scripts in Azure Databricks using PySpark to create linkages between 40 MM doctors and their affiliated organizations, to provide detailed insights to the health insurance companies at an organizational level

***Data Science Intern*** – Divergence.ai, Dallas, TX **June 2019 – August 2019**

- 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
- Integrated the bot with the customer database, and LUIS an Azure cognitive service that provides a natural language understanding capability to the chatbot, and built NLP models for topic modeling
- Integrated Power BI with the chatbot to make it return visualizations based on the questions asked by the customers

***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 the data with 30 features by cleaning and transforming it, and performed descriptive analysis using python
- Built a logistic regression model to predict the purchase likelihood of a customer and identified the variables that impacted customer conversion significantly, resulting in an 18% increase in the customer conversion rate

## HACKATHONS AND ACADEMIC PROJECTS

**Informs Analytics Challenge: LA Restaurant Data Analysis** *(Machine Learning, Python, Tableau)* **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%

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

- Implemented a sign language interpreter to help more than 70 MM people across the world, having a condition of hearing impairment, with their daily communication needs using CNN/Deep Learning, utilizing TensorFlow

**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, found best parameters using K-fold Cross Validation and achieved an accuracy of 62%

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

- Built machine learning 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