Satwant Singh

In: https://in.linkedin.com/in/satwant-singh201
Gitfolio: https://satwant-ds.github.io/Satwant-DS

satwants@usc.edu
(213) 326-8089
Los Angeles, CA, 90007

EDUCATION

University of Southern California	Los Angeles, CA
Master of Science in Applied Data Science	**AUG 2021-MAY 2023**

Relevant Coursework: Foundations of Data Management, Machine Learning for Data Science

Punjab Engineering College	Chandigarh, India
Bachelor of Engineering in Electronics & Electrical Communication - **GPA: 9.4/10**	**AUG 2013-MAY 2017**

Achievements: TCS Gold Medal for Best Capstone project 2017 (**Brain Computer Interfacing**)

SKILLS

- **Languages**: Python, Apache Hive, PySpark, SQL, C++, C, Basic HTML & CSS
- **Data Science Skills**: Linear Regression, Random Forest, GBM, Clustering, CHAID Analysis, Word Embeddings, Map Reduce, Spark, Sequence Modeling, H2o, TFIDF, Recommender Systems, Risk Analytics, Hadoop
- **Data Science Libraries**: Pandas, H2o, Lime, NumPy, Scikit-Learn, NLTK, Gensim, Fastai,Keras, Spark
- **Tools**: Jupyter Notebook, MS Azure Cognitive Services, Linux, Git, Docker, MS SQL Server,DynamoDB, Firebase, MS Excel, AWS EC2

WORK EXPERIENCE

UnitedHealth Group – Center for Advanced Research and Analytics	Gurugram, India
Data Scientist (Project Lead)	**AUG 2019-JUL 2021**
UnitedHealth Group – Customer Experience Analytics	Noida, India
Associate Data Scientist	**JUN 2017-JUL 2019**

PROJECTS

Non-Clinical Audit Prioritization Algorithm (Python, SQL, GBM, H2O, Clustering, CHAID)　　　　　**OCT 2020-JUL 2021**
- Designed Inventory prioritization solution with 95% Precision aimed at reducing False Positive Rate of Audit Process
- Reduced FPR for client team by 4% and increased Audit driven Savings by 3M USD per Month by deploying GBM Classifier
- Standardized Audit process enabled Client to lower audit dollar threshold by 20% and helped identify unknown error patterns

Medical Fraud Risk Authentication (PySpark, Apache Hive, GBM, TFIDF, Optum Clinical BERT Model)　　　**NOV 2019-SEP 2020**
- Mined new Error Patterns by building a semi supervised clinical risk prediction model; Automated 2 error categories with 100% True positive rate; Created Propensity model for 4 value streams within 100M USD Payment Integrity Portfolio
- Collaborated with Client to identify new fraud patterns and helped reduce FPR for overall Clinical Investigation by **7%** and enabled reduction in administrative costing of **4M USD per month**

AutoML Toolkit (PySpark, Statistics, Feature Engineering, Python, Spark-SQL)　　　　　**JUN 2019-SEP 2019**
- Orchestrated a framework for creating 4 Feature types in Risk Analytics namely Look-back, Risk, Entropy and Risk Severity
- Scripted stable version of Toolkit used by 13 Data Science Teams across 20+ machine learning projects

Solve My Problem (SQL, Sentiment Analysis, Python, TFIDF, Clustering, ALS)　　　　　**MAR 2018-DEC 2018**
- Identified top 20 highest customer churn rate Insurance plans in Texas, USA for 2017 & recognized 5 major **Root causes**
- Setup collaborative filtering based recommendation engine for creating personalized user profiles for suggesting better plans to churn prone customers; Churn rate reduction by 2.3% in 2019

Business Insights Web Application (SQL, Flask, NLP, D3.js, NLTK, n-gram analysis)　　　　　**JAN 2018-MAR 2018**
- Released a Flask based Web Application capable of generating SQL query & 3 types of visualizations based on natural language

LEADERSHIP & INVOLVEMENT

- Managed Program Management for 2 Associate Data Scientists & coordinated in **Recruitment Panel** in Team
- Organized Analytics Carnivals, Hackathons & Fresher Onboarding for Optum Analytics (1000+ Folks) from 2018 to 2020
- Coordinated 3 Technical Events with 100+ participation for IETE Society PEC Chapter during University Festival 2016
- Led a team responsible for conducting blood donation camps for National Service Scheme PEC Chapter in 2014 & 2015

CERTIFICATIONS

- Machine Learning with Python (Coursera), Neural Networks & Deep learning (Coursera), Sequence Modeling (Coursera)
- Databases & SQL for Data Science (IBM), Data Visualization with Python (IBM), NLP in Tensorflow (Coursera)