SIDDHARTH GOEL

267-279-8986 | sigoel@seas.upenn.edu | linkedin.com/in/siddharth-goel-in | laughbuddha.github.io

EDUCATION

University of Pennsylvania | **MSE in Data Science (December 2021)** | GPA: 3.72/4.00

- **Coursework:** *Machine Learning, Big Data Analytics, Deep Learning for Data Science, Computational Linguistics (NLP), Statistics for data science*

University of Delhi | **BE in Information Technology (May 2012)** | GPA: 7.75/10

- **Coursework:** *Programming Languages, Object Oriented Design, Algorithms and Data Structures, Database Systems*

KEY SKILLS

- Python – PyTorch, NumPy, pandas, scikit-learn, matplotlib, seaborn
- SQL, NoSQL, MongoDB, Apache Spark

INTERNSHIP

BharatPe, New Delhi, India | Data Science/Machine Learning Intern **Jul 2020 – Dec 2020**

#Series_C_Indian_Fintech_Startup, #Digital_Payments, #5.5M_merchants, #45M_monthly_transactions

- **Predicting Merchant Business Category** | *Python, pandas, scikit-learn, SQL*
  Developed a predictive model to identify the merchant's business category - food & beverages, fuel, consumer goods etc., contributing to **2% revenue increase** by running targeted marketing campaigns and promotional offers.
- **Identifying Money Leakage** | *Python, pandas, SQL*
  Built and conducted experiments to verify all the incoming and outgoing payments reconciliation for all product lines, proactively **preventing potential bottom line impact** by checking money leakage and future cost in audit disputes.
- **Daily Settlement Ratio** | *Data Mining, Python, pandas, SQL*
  Improved the daily payment settlement ratio by identifying users having dysfunctional accounts for reasons such as blocked/frozen accounts, limit exceeded etc. using pattern recognition in recent historical transactions, **increasing user retention by 5%**.

ACADEMIC PROJECTS

- **US Traffic Accidents** | *Predictive Analytics, Python, scikit-learn, matplotlib, seaborn, Spark* | [Github Repo] **Apr 2020**
  Built the complete data science pipeline by performing extensive exploratory data analysis, data pre-processing, feature engineering, and data modelling on about 3 million records of the US Traffic Accidents dataset.
- **Audio Source Separation** | *Deep Learning, PyTorch, Python* | [Github Repo] **Apr 2020**
  Separated MUSDB18 dataset mixture tracks into vocals, drums, bass, and other instruments using LSTM and state-of-the-art deep learning models.

PROFESSIONAL EXPERIENCE

June 2012 – August 2019

Deloitte Consulting (US-India) | Technology Consultant

- **Preventing payment disputes:** Built a classification model using random forest algorithm to predict payment disputes on open invoices, decreasing the effort in consolidation of payments at quarter end by 30% and **reducing bottom-line impact by 25%**.
- **Inventory replenishment**: Developed novel solution to forecast demand from customers' buying trends using croston algorithm and SAP Predictive Analytics Library (PAL), **reducing material stock-outs by 95%** and adding 15% to top-line.
  - The ML solution is estimated to generate revenues worth $10M+ for Deloitte.
- **Demand prediction:** Developed time series ML model using arima algorithm to predict production order time for a manufacturing client. The prediction results helped in confirming delivery time to customer in material routing, leading to **accurate estimation of order delivery** and **reduced gap between planned and actual**.
- **SAP Invoice Management:** Implemented Opentext Vendor Invoice Management (VIM) framework to **create and process ~15,000 vendor invoices monthly**, for a multinational professional services firm.
  - Developed end-to-end invoice handling process including reading of pdf/.doc(x) documents using OCR, exception handling, invoice approval workflow, and invoice posting and storage in SAP system.