APURV PRIYAM

Junior Data Scientist

+1 404-483-5425

apurvpriyam@gmail.com

https://apurv.page

Atlanta, GA

CORE SKILLS

- **Languages:** Python, R, SQL, MATLAB
- **Machine Learning Libraries** TensorFlow, Keras, H2O, Scikit-learn, XGBoost
- **Visualization / Dashboard:** R Shiny, Flask, CSS, HTML, Tableau, D3, ggplot2, Plotly
- **Big Data and Others:** Microsoft Excel, git, Hadoop, MapReduce, MySQL

AWARDS

- Won **ZS Innovator** of Month for Attribution Modeling App that automated the marketing mix projects
- Won **ZS Hackathon** for app that detects anomaly in heartbeat using **Fitbit**'s live data and sends real-time alerts
- **Amgen Inc. Hackathon** - Competed against **5** companies and secured a project for ZS
- Won **ZS best project of the year** for a product that helps sales representatives in efficient customer targeting *(created data pipelines and analyzed data with **2B+** rows using **SQL** and **Hive**)*
- **National Level Data Science Challenge** - Ranked **1st** among 5000+ participants by predicting if customers will buy

EDUCATION

Georgia Institute of Technology, GA, US — **Aug 2019 - Dec 2020**

Master of Science in Computational Science | CGPA: **3.85** / 4

Courses *– Machine Learning, NLP, Deep Learning, Regression, Database, Algorithms*

Graduate Teaching Assistant *– Lectured 150+ students on Machine Learning; prepared and graded assignments*

Indian Institute of Technology, Kharagpur, India — **Aug 2012 - Jul 2016**

Bachelor of Technology in Industrial Engineering | CGPA: **8.17** / 10

WORK EXPERIENCE (3.5 years)

Summer Intern | WESTERN DIGITAL — **Jun 2020 - Jul 2020**

- **Information extraction from invoices using NLP -** *Reduced invoice auditing time by **90%***
  - Trained TensorFlow **Bi-LSTM** model for information (entity) recognition (**0.93** F1)
  - Created word embedding by concatenating **GloVe** and character embedding
  - Developed **Flask** application to load invoices, extract, and correct information in batches
- **Contract clauses review using NLP–** *Automated review of **1000+** contracts / year*
  - Built stack of models to identify risky clauses and then classify them into 8 types (**0.91** F1)
- **Damage detection –** *Expected to save **$0.5M /** year and speed up damage claims process*
  - Created image dataset and trained Computer Vision model to detect damages in shipment

Data Scientist | ZS ASSOCIATES — **Jul 2016 - Jul 2019**

- Among top **10%** of the batch to receive promotion in 4 semesters
- Initiated and led a program to automate common tasks by creating libraries and apps

Key projects

- **Next Best Action** - *Increased sales by **4% ($30M)** and marketing engagement by **25%***
  - Optimized sequence of marketing channels using **Genetic Algorithm (GA)** and **Python**
  - Optimization was personalized for each customer and subjected to business constraints
  - Trained Deep Learning **(CNN)** model to predict sales of sequences (GA's fitness function)
  - Applied **Collaborative Filtering** to find customer's affinity for each marketing channel
- **Attribution Modeling App -** *Cut all project's time by **2 weeks**; acquired new business (**$1M+**)*
  - Wrote **R package** to automate marketing mix project involving data with **50M+** rows
  - Built **R Shiny** Web Application for business users to manage and run projects efficiently

Other Projects

Extensively used Linear and Logistic Regression, Clustering, Random Forest, Gradient Boosting

- Quantified impact of marketing channels on sales and optimized spend leading to **$20M** profit
- Identified geo-segments based on sales growth and found key drivers using **Decision Tree**

R Instructor (Part-time) | WILEY PUBLISHING, INC. — **Apr 2019 - May 2019**

- Created video tutorials on 'Data Analytics with R' covering data processing and modeling

Summer Intern | TATA CONSULTANCY SERVICE — **May 2015 - Jun 2015**

- Developed a simulator in **MATLAB** to generate 72 metrics of IT Service Management
- Analyzed metrics using **Factor Analysis** to build a Decision Support System

PROJECTS

- Built **recommendation system** using **regularized SVD,** and **Collaborative filtering (0.94 RMSE)**. Designed **incremental** method for new ratings - For **250** (15%) new movies, incremental training took **4%** of original time with similar RMSE**.** Implemented SVD function using **Jacobi Transformation** and matrix **Bidiagonalization** using **Householder** matrix
- Published R package **'analyzer'** on **CRAN** (official R repository). It automatically generates interactive notebook with pre-written codes to analyze data using plots, **15+** statistical and hypothesis tests, variable selection, and models
- Created app to find and visualize safest travel path utilizing **Dijkstra's** algorithm. It generates alerts when the driver is near accident zones (using **KDE**). Integrated GPS for navigation
- **Travelling Salesman Problem** using Genetic Algorithm - coded from scratch and compared different local search algorithms, selection process, mutation, and crossover operators