APURV PRIYAM

Advance Analytics, Data Science, Machine Learning | 3.5 Years' Experience

+1 404 483 5425 | apurvpriyam@gmail.com | https://apurv.page | Atlanta, GA

EDUCATION

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

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

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

Bachelor of Technology (Hons.) in Industrial Engineering | CGPA: **8.17**/10

WORK EXPERIENCE

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

- **Information extraction from invoices using NLP -** *Reduced invoice auditing time by **~90%***
  - Developed TensorFlow **Bi-LSTM** model for entity recognition (**0.93** F1); Used **Tesseract** & **OpenCV** for OCR
  - Created word embedding by concatenating **GloVe** embedding & outputs from another **Bi-LSTM** layer trained on **character embedding** of the word's characters
  - Created **Flask** application as interface to load invoices in batch, extract, check and update information
- **Automated contract review using NLP –** *Expected to help in automated reviewing of **1000+** contracts / year*
  - Built layer of models to first identify risky clauses using **BERT** (**0.84** F1); then classify it into 8 types (**0.91** F1)
- **Damage detection using Computer Vision –** *Expected to save **$0.5M** / year and speed up damage claims process*
  - Created custom dataset using shipment images and trained **YOLO** for detecting damages in shipment boxes

Graduate Teaching Assistant | GEORGIA INSTITUTE OF TECHNOLOGY **Jan 2020 - Present**

- Taught and guided **150+** students on Machine Learning and Data Analytics; Prepared and graded assignments

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

Key projects

- **Next Best Action –** *Increased sales by **4% ($30M)** and digital marketing channel engagement by **25%***
  - Optimized sequence of marketing channels, personalized for each customer and subject to business constraints
  - Used **Genetic Algorithm** with Deep Learning (Tensorflow **CNN+Dense layer**) model as its fitness function
  - Applied **Collaborative Filtering** to find customer's affinity for each marketing channel (email, mobile, push etc.)
- **Attribution Modeling App –** *Cut down project timelines by **1-2 weeks**; Acquired new projects and clients **($1M+)***
  - Wrote **R package** to automate data processing, exploration, and modeling for promotional response projects
  - Created **R Shiny** Web Application as user interface for business users to manage and run projects efficiently

Other Projects

- Calculated impact of marketing channels on sales using **GLM** & optimized spend leading to **$20M** profit for clients
- Quantified effects of marketing activity on patient transitions & product sales using **Markov Chain** and **GLM**
- Identified different geo-segments based on sales growth and found responsible key drivers using **Decision Tree**

Other Activities

- Among top **10%** of the batch to receive promotion in 4 semesters
- Initiated and led a program named 'Automation' to automate tasks and reduce rework by creating libraries and products
- Trained colleagues, including seniors, on R programming, Genetic Algorithm, and Machine Learning

R Programming Instructor | WILEY INDIA PVT. LTD. **Apr 2019 - May 2019**

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

Summer Intern | TATA RESEARCH DEVELOPMENT & DESIGN CENTRE **May 2015 - Jun 2015**

- Developed a data simulator in **MATLAB** to generate 72 metrics under IT Service Management
- Analyzed metrics data using **Factor Analysis** to build a Decision Support System for better utilization of resources

PROJECTS

- **Recommender System using low-rank approximation -** github.com/apxr/RecSystem **2020**
  - Built a recommender system in **MATLAB** using **SVD, regularized SVD,** and **Collaborative filtering** (RMSE **– 0.94**)

- Designed **incremental** method to handle new ratings without re-calculating full SVD - Re-training after addition of **250 (15%)** new movies took **3.9%** of original time while error increased by **0.2%**

- **'analyzeR' R package** - wrote & published on official R repo - CRAN (cran.r-project.org/package=analyzer) **2020**
  - It automatically generates interactive notebook with pre-written codes to analyze data using 15+ statistical & hypothesis tests, plots, variable selection, and models
- **SPOT** (Safest Path Optimization Tool) - github.com/apxr/SPOT **2019**
  - Created an app to find and see safest travel path utilizing **Dijkstra's** algorithm; Integrated GPS for navigation
  - Used **Kernel Density Estimation** to measure driver's proximity to accident-prone zones. App generates alert when driver is near such zones
- **Travelling Salesman Problem using Genetic Algorithm -** github.com/apxr/TSP_GA **2019**
  - Implemented from scratch in Python. Compared different selection process, mutation & crossover operators

CORE SKILLS

Languages:	Python, R, MATLAB, SQL, C++
Machine Learning:	TensorFlow, Keras, H2O
Web Apps:	R Shiny, Flask, CSS, HTML, D3.js
Others:	MS-Excel, git, Hadoop, Tableau

AWARDS AND HONORS

- Won **ZS Innovator** of Month award for Attribution Modeling App that automated promotional response projects **| 2019**
- **AIMinds** (seminar organized by Analytics India Magazine) - Presented on sequence optimization using Genetic Algorithm and Deep Learning **| 2018**
- Won **ZS Hackathon** for app that detects anomaly in heartbeats using **LSTM (Keras)** model & **Fitbit**'s live data; Sends alerts in real-time (used **R, R Shiny**) **| 2017**
- **Amgen Inc. Hackathon** - Competed against **5** companies and secured project on promotional response for ZS **| 2017**
- **ZS best project of the year** - Won for a product that helps sales representative to target right customer at right time with right content (used **SQL**) **| 2017**
- **National Level Data Science Challenge** - Ranked **1st** among **5000**+ participants by predicting the propensity of a customer to buy food and beverages; Built a predictive stacked ensemble model using 3 models **| 2015**