AMANDEEP RATHEE

amandeep.rathee@duke.edu
amandeeprathee.com

Data Scientist
AWS Cloud Certified ● Kaggle Top 1%

Work Authorization: F1-OPT
Durham, NC

EDUCATION

- **Duke University, Durham, NC**
  Master of Science in Interdisciplinary Data Science; GPA: 3.8/4.0 *August 2019 – May 2021*
  Teaching assistant for two graduate courses: *Data Visualization* and *Healthcare Analytics*.

- **Visvesvaraya Technological University, Bengaluru, India**
  Bachelor of Engineering in Computer Science and Engineering *August 2013 – June 2017*

RELEVANT WORK EXPERIENCE

- **Machine Learning Intern, LifeOmic Inc.** Morrisville, NC; May 2020 - August 2020
  - **Medical Document Denoiser**: Reduced the noise removal time of a convolutional neural network by 4x that removes artificial noise from medical documents.
  - **Intelligent Medical Document Router**: Created a linear regression model that saves 80% time in the medical document noise removal process by quantifying the amount of noise present in the documents.
  - **Tools and Technologies**: Python, AWS, Git, CI/CD, Convolutional Neural Networks, U-Net, OCR, Quantization.

- **Data Scientist Intern, Duke Center for Computational Biology** Durham, NC; November 2019 - May 2020
  - **RPhenoscape**: Developed a feature for an open-source R package that computes mutual exclusivity among various phenotypes from different species based on information extracted from unstructured ontology-based data.
  - **Tools** R, Git, Web Ontology Language (OWL).

- **Content Strategist - Data Science, UpGrad Education** Bengaluru, India; July 2017 - May 2019
  - **Telecom Churn**: Created machine learning models for predicting and making inferences about customer churn for one of India's largest telecom operators that helped the company make more targeted ads to retain customers and save 80% in customer acquisition.
  - **Data Science, Machine Learning & AI Graduate Programs**: Developed modules on data science, machine learning, deep learning, and natural language processing that increased revenue of these one-year online programs by $500,000, drove the net promoter score from +25% to +60%, and helped them become top online AI programs in India.
  - **Courses and Modules Created**: Healthcare Analytics, Natural Language Processing, Neural Networks.

SKILLS

- **Supervised Machine Learning**: Regression, decision trees, random forest, gradient boosting machines, support vector machines, k-nearest neighbors, neural networks, transfer learning, ensemble methods, model stacking, and quantization.
- **Unsupervised Machine Learning**: K-means clustering, hierarchical clustering, and principal component analysis.
- **Reinforcement Learning**: Policy-based learning, value-based learning, actor-critic methods, and trust region policy optimization.
- **Experience in**: statistical analysis, machine learning, deep learning, reinforcement learning, natural language processing, computer vision, data analysis, and data visualization.
- **Languages**: Python, R, Java, C++, and SQL.
- **Tools**: Pytorch, tensorflow, keras, scikit-learn, pandas, numpy, matplotlib, tidyverse, and Tableau.
- **Software Engineering and Cloud Computing**: AWS Certified Cloud Practitioner, software design patterns, SCRUM and agile development, CI/CD pipeline, Git, and CLI.

PROJECTS AND ACHIEVEMENTS

- **[Graduate Capstone] Solving Vehicle Routing Problem Using Reinforcement Learning:** Coming up with heuristics using attention-based reinforcement learning methods to solve a NP-hard vehicle routing problem.
- **[Undergraduate Capstone] Helmet Detection on Two-wheeler Riders Using Machine Learning:** Trained an ML model that detects riders riding a two-wheeler without a helmet. Also published as a paper in the IJAECS journal.
- **Playstation vs Xbox:** Conducted statistical analysis to determine that PlayStation games have 30% higher sales than Xbox games.
- **Sentiment Analysis:** Analyzed public opinion on Demonetization – a highly controversial currency ban that happened in India in 2016 – using Twitter API, and found that 50% of the tweets were neutral, 30% were in favor and 20% were against the ban.
- **Kaggle Top 1%:** Ranked #33 among 150,000+ data science practitioners on Kaggle as a result of winning one gold, one silver, and seven bronze medals by sharing insightful analyses using open datasets.