navyad265@gmail.com
nd265.github.io

NAVYA DAHIYA

+1 (778)318-3700
Vancouver, Canada

Education

Masters in Data Science, *University of British Columbia, Vancouver, Canada* **Sep 21–Jun 22**
Bachelors of Technology in Computer Science & Engineering, *Shiv Nadar University, India* **Jul 15–May 19**

Skills

Programming Languages	Python, R, Java
Databases	MySQL, MongoDB, Postgres
Data Science toolkit	Computer Vision, Machine Learning, Deep Learning, PyTorch, NLP, Transformers, Data Visualization, Experimentation and Causal Inference, A/B Testing, Hypothesis Testing, Statistical analysis
Big Data toolkit	Hadoop, Hive, Spark, Flume, AWS - S3, EC2, IAM
Software Development	Docker, CI/CD pipeline, Full Stack Development

Academic Projects

Head Collision Detection in Sports Videos **May22-Present.**
- Implementing SSD, YOLO and 3DCNN algorithms to detect occurrence of head collisions in sports videos, which would save 90% of the time spent manually in the analysis and detection of collisions in the entire game

Rainfall Prediction **Apr 22**
- Predicted daily rainfall in Australia using Random Forest Regressor and deployed the ML pipeline using Flask API with Spark instance in AWS

MindtheGap, MindtheGapR **Mar 22**
- Created interactive dashboards for analysing Gapminder dataset using Dash, python, R and deployed the apps on Heroku

Simplefit, Simplerfit **Feb 22**
- Created Python and R packages which fit and compare regression, classification ML models' scores with baseline models and return EDA plots, thus reducing 80% of a data scientist's effort. Built and deployed the packages using CI/CD pipelines.

Census Income Prediction **Dec 21**
- Predicted annual income from census data having demographic features using Random Forest Classifier and deployed the pipeline on Docker.

Technical Experience

SOFTWARE ENGINEER 2 | Dell International Services, India **Jul 19–Sep 21**
- Designed and developed an in-house performance testing tool from scratch that enables QAs to schedule load tests, identifies performance bottlenecks in Dell web applications, shares customized test reports via email and analyses the quality of the test, thereby reducing time consumed in operational overhead by 40%. The tool is used by 65% of engineering projects in Dell
- Led a team of 3 members and integrated the tool with CI/CD pipeline, which reduced the number of production failures by 15% over the course of 18 months

BIG DATA INTERN | Birlasoft Pvt Ltd, India **May 18–Jul 18**
- Performed Twitter Sentiment Analysis on FIFA World Cup tweets using Hive, Flume and Python
- Built a java library that is integrated in ETL pipelines across the team and is capable of migrating more than 1000 schemas from Oracle DB into a data warehouse (Hive) by bridging the gap between their schemas' data types

Achievements

Women in Data Science Hackathon | Kaggle **Mar 22**
- Ranked 1st in Vancouver and top 2% out of 829 (16/829) teams worldwide
- Used ensemble modelling along with PCA and clustering techniques to predict the building energy consumption

Bot Detection model | Dell AI/ML Hackathon, India **Aug 19**
- Ranked 6th among 40 teams in Dell
- Used Naïve Bayes algorithm to classify the mouse movements as Bot or Human on Dell's home page, based on mouse mapping

Prediction of MSP for Farmers | Dell Recruitment Hackathon Drive **Aug 18**
- Predicted Minimum Support Price for the Indian Farmers using Linear Regression(0.83 R2score) which resulted in a job offer by Dell International Services

EXTRA CURRICULARS

- Volunteer of Content team in DataCan, a Women in Data Science community **Dec 21–Present**
- Graduate Block 2 Representative, University of British Columbia, Vancouver **Oct 21–Nov 21**
- Chairperson of Academic Affairs Committee, Shiv Nadar University, India **Jan 19–Jan 20**