

Dipit Vasdev

New York, NY | [\(347\) 699-9484](tel:(347)699-9484) | dipit.vasdev@nyu.edu | [linkedin.com/in/dipit-vasdev](https://www.linkedin.com/in/dipit-vasdev) | github.com/dipitvasdev

EDUCATION

New York University – New York, NY

Master of Science | Computer Engineering – GPA: 3.9 / 4

September 2022 – Present

Graduation: 15th May 2024

Relevant Coursework: Machine Learning, Data Structures and Algorithms, Data Science, Deep Learning, Data Visualization, Big Data

Relevant Skill from Coursework and Projects:

- **Introduction to Spark for Big Data:** Acquired skills in utilizing Apache Spark for efficient data processing and analysis, focusing on resilient distributed datasets (RDDs), transformations, and actions to construct optimized data pipelines
- **Distributed Machine Learning with Spark:** Studied the application of Spark in distributed machine learning, emphasizing iterative algorithms and optimization techniques to enhance computational efficiency for large-scale data analysis.

Guru Gobind Singh Indraprastha University – New Delhi, India

August 2018 – July 2022

Bachelor of Technology | Computer Science and Engineering – CGPA: 9.25/10

Research Publications: Vasdev, D, Jain, M, Pal, K, Sharma, V. "Systematic literature review on predictive maintenance of vehicles and diagnosis of vehicle's health using machine learning techniques". Computational Intelligence. 2022; 1- 19. doi:10.1111/coin.12553

SKILLS

Core Skills:	Machine Learning, Deep Learning, Data Science, Natural Language Processing, Computer Vision
Big Data and Cloud:	Relational Database, Generative AI, Apache Hadoop, Apache Spark, Dash, Google Cloud Platform
Programming Languages:	Python, C++, Kotlin, Java, SQL, TypeScript, JavaScript
Data Analysis and Visualization:	Pandas, NumPy, Matplotlib, Power BI, Tableau, Microsoft Excel
Frameworks and Libraries:	TensorFlow, Flask, Scikit-learn, Git, Node.js, Express.js, MongoDB, Mongoose, PyTorch, React

EXPERIENCE

Graduate Course Assistant | New York University (NYU), New York, NY

September 2023 – Present

- Collaborated with other TAs to facilitate all aspects of graduate-level Deep Learning course for 100+ students, including grading, assignments, and doubt resolution.

Software Engineer Intern (Gen AI) | Alten Capital, New York, NY

May 2023 – August 2023

- Developed highly scalable backend systems using leading technologies, leading to a 17% reduction in redundant code.
- Leveraged OpenAI function calling API, langchain and zero/few-shot learning to enhance conversational capabilities.
- Developed expertise in prompt engineering, including techniques for formulating effective prompts, evaluating model performance, and mitigating bias.

Software Engineer Intern | S&P Global (Standard & Poor's), India

June – July 2021

- Extracted more than 10,000 ALB Access logs/day from AWS S3 bucket using regular expressions and AWS Query Language
- Reduced infrastructure costs by 23% by leveraging AWS Lambda and CloudWatch APIs to streamline data extraction
- Configured Logstash to send AWS Cloud RDS data to Elasticsearch, facilitating comprehensive data analysis

PROJECTS

NYU Data Science Capstone: Exploring Music Data Analytics (Pandas, SciPy, Python, Scikit-learn)

December 2023

- Led a data analysis project in the music domain, leveraging statistical tests (U test, KS test, Cohen's D) and various machine learning algorithms (OLS, Ridge, Lasso, Random Forest, Decision Trees) to drive robust data insights.
- Utilized PCA with Kaiser Criterion to identify 3 key components explaining over 57% of data variance and determined optimal k-means clustering (k=4) via silhouette scores to improve data segmentation.
- Implemented collaborative filtering with Alternating Least Squares (ALS) in a recommender system for music recommendation

NYU Deep Learning: Spoiler Detection in Movie Reviews using Distil BERT (Python, NLP, Flask)

May 2023

- Developed and implemented deep learning techniques for spoiler detection in movie reviews, achieving 78.49% accuracy with Distil BERT which beats competing research work on that dataset.
- Conducted thorough data pre-processing on a 1 GB dataset with 573,913 reviews, removing URLs, converting text to lowercase, eliminating special characters, and handling stop words for improved model performance.

AWARDS AND ACCOLADES

Recognized as Top Data Science Voice – LinkedIn

January 2024

Runner up, eYantra Robotics Competition – Indian Institute of Technology, Bombay

April 2021