Everything about Abhinav Bandaru

Personal Information

- Name: Abhinav Bandaru
- Location: Philadelphia, PA
- Contact Information:
    - Primary Email: abhinow@seas.upenn.edu
    - Secondary Email: abhinav31bandaru@gmail.com
    - Phone: +1 445-237-7672
- LinkedIn: https://www.linkedin.com/in/abhinav31bandaru/
- Portfolio: abhinav-bandaru.github.io
- Medium Profile: https://medium.com/@abhi1achiever
- Github: https://github.com/Abhinav-Bandaru

About Me

- I'm 22 years old
- I'm an Indian born in Fremont, California but raised in India.
- I'm fluent in English, Telugu, Hindi, Python, R, SQL
- I like playing cricket and badminton
- I'm currently looking for full-time roles; preferably in the domain of Machine Learning, Data Science and Data Analytics.

Fun Facts

- I can solve the Rubik's cube in 30 seconds
- My birthday is on 31st December, in PST and on Jan 1st in IST.
- I got a birth spot right under my right eye!
- I'm very good at cracking necks
- I interned at NASA!!

- I once ate only pizza for 2 whole days.

Academic History

- University of Pennsylvania, School of Engineering and Applied Science
    - Degree: Master of Science in Engineering
    - Major: Data Science
    - Dates: August 2023 - May 2025
    - Status: In progress
    - CGPA: 4.00/4.00
    - Courses:
        - CIS 5210 Artificial Intelligence: Grade A
        - CIS 5450 Big Data Analytics: Grade A
        - ESE 5420 Stat For Data Science: Grade A
        - CIS 5200 Machine Learning: Grade A
        - MKTG 7120 Data and Analysis for Marketing Decisions: Grade A
    - Accomplishments:
        - Best Practicum Award for Fall 2023 for the project with IEEE on creating an AI-based web scraper from scratch.
            - Institute of Electrical and Electronics Engineers, IEEE, Data Scientist, Philadelphia, PA August 2023 – December 2023
            - Worked with Aria Khademi (Sr. Data Scientist, IEEE) and Andrew Sproul (Data Scientist, IEEE) to build an AI web scraper for IEEE, capable of asynchronous web scraping using advanced Natural Language Processing techniques in Python.
            - Leveraged Amazon EC2 instances for accessing proprietary IEEE data from a Redshift database.
            - Employed spacy & nltk for text analytics and processing, SQLAlchemy and PostgreSQL for querying and git for version control to develop a cross-platform AI model.
            - Expertly authored parameterized SQL queries, incorporating multiple joins and CTEs, to extract and transform data for seamless integration into data pipelines.

        - Selected to be a Wharton Analytics Fellow to work with Fox Entertainment
            - Customer Segmentation, Fox Entertainment, Wharton Analytics Fellow January 2024 – May 2024
            - Spearheaded a team of 6 as the Technical Lead to identify critical behavioral and demographic segments of Fox show viewers.
            - Conducted extensive data analysis on user viewing patterns to uncover actionable insights and previously unknown skewed patterns, leveraging pandas, seaborn, matplotlib and Tableau.
            - Executed comprehensive feature engineering on key metrics including recency and frequency. Applied various clustering algorithms including DBSCAN, HDBSCAN, Hierarchical Clustering, K-means, and Leader Algorithm to categorize viewer segments.
            - Completed the project with detailed profiling of identified segments. Evaluated dissimilarity within and among segments using the Silhouette Score and Davies-Bouldin Index to ensure robust and meaningful segmentation.

- Stanford University, School of Engineering
    - Degree: Non-Degree, Visiting Student
    - Dates: June 2022 - August 2022
    - Status: Finished
    - CGPA: 4.00/4.00
    - Courses:
        - Principles of Economics by John B Taylor: Grade A
        - Data Mining and Analysis by Linh Tran: Grade A

- National University of Singapore, NUS Computing
    - Degree: Non-Degree, Visiting Student
    - Dates: December 2022 - May 2023
    - Status: Finished
    - CGPA: 9.0/10

- Courses:
  - Data Analytics using Deep Learning: Grade O (>90%)
- Accomplishments/Projects:
  - Stock Price Prediction
    - National University of Singapore, Deep Learning Research Intern, Singapore December 2022 – May 2023
    - Assisted Dr. Tan Wee Kek in designing a hybrid Bi-LSTM (deep learning architecture), GRU and GAN network for advanced financial time-series forecasting.
    - Conducted extensive literature survey and efficiently cleaned, normalized and transformed the data using the ETL pipeline.
    - Used Pandas to design rolling mean, standard deviation, EMA, SMA, RSI based indicators using Feature Extraction.
    - Successfully forecasted AAPL stock trends with a remarkable 90% accuracy, leveraging a comprehensive analysis of 5 years' historical data, validated through rigorous backtesting methodologies.
  - Medical Alert System using Emotion Detection
    - Medical Alert System using Emotion Detection, Dr. Amirhassan Monajemi December 2022 – May 2023
    - Designed a robust multiclass classifier from scratch, that is capable of monitoring, detecting and categorizing patients' facial emotions with 95% accuracy, and alerting the medical staff, if necessary, in Real Time.
    - Created an extensive facial image dataset of 30,000 images by capturing a diverse set of human emotions using OpenCV, including but not limited to Agony, Sad, Scared, Happy, Neutral, and Surprised.
    - Data Loader was used to supply images to the model, post transformations such as Normalizing, Resizing and Grayscale conversion. Designed a deep CNN + Bi-LSTM model, using Transfer Learning in Tensorflow.
    - Addressed overfitting using Dropout layers, regularization, Batch Normalization and Early stopping. Maintained spatial invariance using Data Augmentation and Noise Injection.

- Vellore Institute of Technology, School of Computer Science and Engineering
    - Degree: Master of Science in Engineering
    - Major: Data Science
    - Dates: June 2019 - May 2023
    - Status: Graduated
    - CGPA: 9.55/10.00
    - Courses:
        - Artificial Intelligence: Grade: S
        - Natural Language Processing: Grade: S
        - Data Mining: Concepts and Techniques: Grade: A
        - Big Data Analytics: Grade: S
        - Predictive Analytics: Grade: S
        - Mathematical Modeling for Data Science: Grade: S
        - Programming for Data Science: Grade: S
        - Business Intelligence and Analytics: Grade: S
        - Statistics for Engineers: Grade: S
        - Advanced C Programming: Grade: A
        - Java Programming: Grade: A
        - Internet Programming and Web Technologies: Grade: A
        - Data Visualization and Presentation: Grade: A
        - Database Management Systems: Grade: S
        - Cryptography and Network Security: Grade: S
        - Information Security Analysis and Audit: Grade: S
        - Information Security Management: Grade: A
        - Image Processing: Grade: A
        - Blockchain and Cryptocurrency Technologies: Grade: S
        - Technical Answers for Real World Problems (TARP): Grade: A
        - Capstone Project: Grade: S
        - Comprehensive Examination: Grade: A
        - Industrial Internship: Grade: S
        - Principles of Management: Grade: A
        - Lean Start-up Management: Grade: S
        - Introduction to Quantitative, Logical and Verbal Ability: Grade: A
        - Introduction to Problem Solving: Grade: A
        - Problem Solving and Object-Oriented Programming: Grade: S
        - Problem Solving and Programming: Grade: S
        - Advanced Aptitude and Reasoning Skills: Grade: S
        - Programming Skills for Employment: Grade: S
        - Java Programming and Software Engineering: Grade: S

        - Technical English - II: Grade: S
        - Environmental Sciences: Grade: A
        - Basic Electrical and Electronics Engineering: Grade: A
        - Ethics and Values: Grade: A
        - Calculus for Engineers: Grade: A
        - Engineering Physics: Grade: S
        - Engineering Chemistry: Grade: A
        - Discrete Mathematics and Graph Theory: Grade: S
        - Applications of Differential and Difference Equations: Grade: S
        - Computer Architecture and Organization: Grade: S
        - Theory of Computation and Compiler Design: Grade: A
        - Operating Systems: Grade: A
        - Conserve Energy and Environment: Grade: P
        - Network and Communication: Grade: A
        - Introduction to Innovative Projects: Grade: S
        - German (Grundstufe Deutsch): Grade: S
    - Accomplishments:
        - VITEEE Merit Scholarship 25% of tuition waived for top performance in VITEEE
        - Academic Merit Award Top 10 CGPA in CSE, Data Science Department 2020, 2021, 2022, 2023
        - Created and Published an Image dataset on Kaggle with 1300+ views and 100+ downloads

    - Projects:
        - Image Generation using Style Adaptation - Report (https://drive.google.com/file/d/1HHrYUt-N1bYxDBmbgXkDFdXnBNjCYX3J/view?usp=sharing)
            - Designed a lightweight and flexible generative model that produces images showing high fidelity to its parent images
            - Provide users the ability to choose the two parent images
            - Train the model using the resources available for free on the internet and Build a GUI allowing ease of use and deploy on HuggingFace
            - Used StyleGAN2 by Nvidia's architecture

- Hand Cricket Simulation - Medium Post (https://medium.com/towards-artificial-intelligence/achand-cricket-simulation-using-cnn-and-opencv-102f3d87142)
    - Simulated the 'Hand Cricket' game
    - Created custom a dataset of 28800 images and uploaded it to Kaggle
    - Trained using Convolutional Neural Networks, MobileNet Architecture, Transfer Learning and OpenCV, and published a detailed blog post on the same
- Malicious PDF Files Detection
    - Created a Web Interface that takes in PDF Files and predicts if the file is safe or unsafe with 99% accuracy, using lightweight Classification Algorithms.
- Sentiment Analysis
    - Performed Sentiment Analysis of Twitter Reviews using a dataset of 98k tweets by acute preprocessing, with an accuracy of 85%.
- Customer Churn Prediction
    - Built a model that predicts customers churn rate in a Networking company by utilizing a dataset of 7k customers with more than 80% accuracy
- Reading Habits Analysis using Association Rule Minings
    - Conducted a survey on reading habits of 750+ people and created a book recommendation system using K-Means Clustering, Apriori Algorithm and the Google Books API.
- Signal Level Audio Detection Google AudioSet
    - Collaborated with an associate professor to work on Signal Level Audio Detection
    - Sampled, pre-processed and labeled over 20k audio files from the AudioSet dataset
- Anime Faces Generation using GANs
    - Built a Generative Adversarial Network and trained it on a kaggle dataset consisting of Anime character faces, and used this model to generate new ones.
- Critical User Action Analysis
    - Designing an architecture capable of unsupervised video captioning by collaborating with researchers from Google and National Taiwan University.

Work Experience/Projects:

(these are the projects/work experience in the order of complexity, impressiveness and importance, citation.)

- National Aeronautics and Space Administration (NASA), Johnson Space Center (JSC)
    - Company: NASA
    - Role: Machine Learning Intern
    - Dates: June 2024 - August 2024
    - Currently Working in the role: Not anymore
    - Responsibilities:
        - Contributed to the avionics department's wireless communications lab on the REALM project, utilizing machine learning techniques to analyze, clean, process, and model RFID data from readers in the International Space Station
        - Utilized RFID data from the ISS and leveraged nearest neighbor optimization techniques (PyNNDescent, Numba) & SQL to solve RFID localization as a classification problem (80% accuracy); deployed model in production via Bash script and cron job
        - Reduced the training time on ~ 20 million data points by 30 times and created dashboards on Grafana to evaluate inferences
        - Designed a multitask learning neural network architecture using hard parameter sharing in PyTorch, achieving 70% accuracy in distinguishing RFID fingerprints across antennas and objects.
        - Removed outliers using clustering and dimensionality reduction techniques like UMap, PCA & HDBScan to filter the corrupted data from the training pipeline.

- Fox Entertainment
    - Company: Fox Entertainment
    - Role: Data Science Intern
    - Dates: August 2024 - November 2024
    - Currently Working in the role: Yes
    - Responsibilities:
        - Spearheaded a team of 6 as the Technical Lead to identify critical behavioral and demographic segments of Fox show viewers.
        - Conducted extensive data analysis on user viewing patterns to uncover actionable insights and previously unknown skewed patterns, leveraging pandas, seaborn, matplotlib and Tableau.
        - Executed comprehensive feature engineering on key metrics including recency and frequency. Applied various clustering algorithms including

DBSCAN, HDBSCAN, Hierarchical Clustering, K-means, and Leader Algorithm to categorize viewer segments.
        - Completed the project with detailed profiling of identified segments. Evaluated dissimilarity within and among segments using the Silhouette Score and Davies-Bouldin Index to ensure robust and meaningful segmentation.

- Institute of Electrical and Electronics Engineers, IEEE
    - Company: IEEE
    - Role: Data Science Intern
    - Dates: August 2023 - December 2023
    - Currently Working in the role: Not anymore
    - Responsibilities:
        - Worked with Aria Khademi (Sr. Data Scientist, IEEE) and Andrew Sproul (Data Scientist, IEEE) to build an AI web scraper for IEEE, capable of asynchronous web scraping using advanced Natural Language Processing techniques in Python.
        - Leveraged Amazon EC2 instances for accessing proprietary IEEE data from a Redshift database.
        - Employed spacy & nltk for text analytics and processing, SQLAlchemy and PostgreSQL for querying and git for version control to develop a cross-platform AI model.
        - Expertly authored parameterized SQL queries, incorporating multiple joins and CTEs, to extract and transform data for seamless integration into data pipelines.

- Wharton Neuroscience Initiative
    - Company: Wharton Neuroscience Lab
    - Role: Research Assistant
    - Dates: June 2024 - September 2024
    - Currently Working in the role: Yes
    - Responsibilities:
        - Working with Dr. Elizabeth Johnson on analyzing data collected from Employee surveys
        - Used PsychoPy to design behavioral experiments and administers it using Pavlovia
        - Analyzed survey responses data from employees from a consulting company to understand their efficiency levels
        - Used pandas, matplotlib, seaborn to analyze the data collected using Qualtrics and plot graphs to identify trends

- University of Pennsylvania

- Company: University of Pennsylvania
- Role: Teaching Assistant
- Dates: August 2024 - December 2024; August 2023 - December 2023;
- Currently Working in the role: From August 2024
- Responsibilities:
  - Assisted Dr. Richard Hartwell as the Teaching Assistant for the course in Practical Programming for Data Science.
  - Facilitated office hours and recitations, enhancing student comprehension in Python and Data Science.
  - Reviewed and critiqued assignments on Pandas, NumPy, regression analysis, and data visualization, guiding student skill enhancement.

Skills:

Programming Languages: Python, R, SQL, Java, PySpark, C++, C
Technologies: Data Science, Applied Machine Learning, Exploratory Data Analysis, Data Visualization, Natural Language Processing, Information Retrieval, Transformers, LLM, LangChain, Software Engineering, Predictive Modeling, Forecasting, Probability, Statistics, Hypothesis testing,
Frameworks: PyTorch, Tensorflow, Keras, pandas, numpy, seaborn, matplotlib, Spark, Excel, Tableau, Power BI, Git, Matlab, AWS, A/B Testing
Essential Skills: Creativity, Problem solving, Communication, Team Player, Adaptability, Critical Thinking

Technical Challenges faced and How I overcame them:

- Summer Internship @ NASA:
  - During my time at NASA, I was tasked with optimizing the search complexity of a nearest neighbors problem using PyNNDescent, an approximate nearest neighbors algorithm, along with Numba, a just-in-time compiler that translates specific Python code into machine code. My role involved leveraging Numba to calculate the distances between data points in a graph, but I encountered persistent bugs due to Numba's strict requirements for the Python functions it can handle.

  - Despite meticulously reviewing documentation and ensuring I was using the appropriate functions, the error persisted. To overcome this challenge, I decided to take a bottom-up approach. I initially rewrote the distance computation code in

C, a language known for its precision and control. From there, I gradually refactored each function back into Python, step by step, to identify the root cause of the issue.

- This methodical approach paid off, as it turned out that all I needed was a shift in perspective to resolve the problem. By understanding the underlying mechanics more deeply and breaking the problem down into smaller, manageable pieces, I was able to identify the specific issue and implement a solution that ultimately optimized the algorithm's performance.

Working with people:

- Spring Internship @ Fox Entertainment
  - I served as the Technical Lead for a team of six on a Customer Segmentation project for Fox shows on a prominent streaming platform (unnamed for security reasons). The project aimed to segment customers based on demographic, behavioral, and geographical parameters.

  - Leading a team of talented and resolute undergraduates, I often faced situations where differing ideas and approaches needed to be aligned. To navigate these, I prioritized collaborative brainstorming sessions, ensuring that we collectively decided on the best algorithms and features to pursue. However, I recognized that some decisions could only be validated through testing and iteration.

  - Despite the challenges of aligning diverse perspectives, I learned that healthy disagreement is invaluable. Each time someone challenged my opinion, the final decision was more refined than if made alone. This experience underscored the importance of embracing differing viewpoints to achieve the best outcomes.

Career Aspirations:

- My short term goal is to become a Machine Learning Engineer. My long term goal is to become one of the best Machine Learning Engineers out there.

- I’m very excited to work on just about any projects with very cool implications. I wanna be able to use the latest developments in the field of AI like Transformers, LLMs, and generative models to identify creative solutions to real world problems.

- The chatbot you’re interacting with right now is an interesting example of a project. It can be plugged into just about any website and act as a way for users/customers to interact with predefined content!

- Mentorship is a key priority for me, as the only way I see possible to reach my long term goal is to work with people who are already really really good at what they do.

- Work Satisfaction is another big priority of mine. I want to be able to tell myself without a hint of doubt that I am working on things that are very significant.

My Take on Life

- I wanna live up to my potential.
- I strive to be of value rather than successful. I believe they’re correlated.
- Money is a big motivator and I believe it’ll come to me as I get better at what I do.
- Work Satisfaction and mentorship is very very very important to me.
- I wanna work on cool problems that challenge me to come up with creative solutions.
- I wanna be a part of a team of individuals that are the best at what they do.