# JANAVI KASERA

Email: jkasera@bu.edu | Mobile: +1 857-264-7093 | LinkedIn: http://www.linkedin.com/in/janavikasera

## EDUCATION

**Boston University**, College of Arts and Sciences                                      *Magna Cum Laude* | GPA: 3.88
*Bachelor of Arts,* Double Major in *Economics* and *Computer Science-Statistics*                    *Graduated May 2023*

## SKILLS

**Programming:** Python (Numpy, Pandas, SciPy, NLTK, Networkx, Seaborn, Matplotlib), Java, C, R, STATA, JMP, MATLAB
**Databases and Visualization:** MySQL, MongoDB, Tableau, Power BI
**Deep Learning Technology:** Tensorflow, PyTorch, Huggingface, OpenCV, LlamaIndex, CNN, RNN, LSTM
**Other:** Microsoft Office (Excel, Powerpoint, Word), Power Apps, AWS
**Publication:** *Hardwiring ViT Patch Selectivity into CNNs Using PatchMixing* ( www.arxiv.org/abs/2306.17848 )

## PROFESSIONAL EXPERIENCE

**Boston University, MIT and Harvard University Joint ML Project**                                      Boston, MA
*Research Assistant in Department of Mathematics and Statistics*                        *August 2022 – Present*
- Lead a team of **3** for developing a user-friendly MATLAB API for **Dr. Mark Kon's**, an associate professor of statistics in Boston University, research algorithm to ensure efficient implementation for other data scientists using sparse and high-dimensional data.
- Coordinate and facilitate weekly lab and research meetings among Professors, post-doctoral researchers, and graduate students across the three universities labs to foster collaboration and exchange of ideas.
- Present and analyze evaluation metrics, such as, confusion matrix, accuracy plots, loss plots, and classification plots to faculty and students to discuss algorithm's performance in data optimization.
- Developing a parameter selection toolkit in MATLAB to enhance the accuracy of predictive modelling datasets.

**AI4ALL – Stanford University's Affiliate**                                      Boston, MA
*Research Assistant and Program Co-ordinator/Organizer*                        *June 2022 – August 2022*
- Developed the **first** ever guidebook on *architecture selection between 2 deep learning models* (CNN and Transformers) while also demonstrated steps to improve selection accuracy by **40%** by analyzing **5+** factorial scenarios for **17+** object classification with **272k+** synthetic images using PyTorch.
- Evaluated the variability of object attributes by creating and pre-processing over **900+** real-life images in Python and applying log-transformations to the images.
- Visualized the test data results, performance comparison and accuracy table in Python with a comprehensive analysis report for the research leads.

**PwC (PriceWaterhouseCoopers)**                                      Gurgaon, India
*Finance Effectiveness Intern in Management Consulting*                        *June 2021 – August 2021*
- Improved **30%** est. managerial efficiency by conducting analysis of **60+** process maps across **9+** financial departments and identifying, eliminating, and consolidating Delegation of Authority points for the internal leadership team.
- Reduced diagnostic turnaround time from **6 months to 1 month** (81%) by developing a prototype of a KPI calculation app built using PowerApps to audit internal finance processes of prospective clients.
- Performed thorough market research and competitive analysis for **4+** clients and presented research reports to Senior Partners.

## PROJECTS & COMPETITIONS

**Feedback Prize-English Language Learning (NLP) Competition** (www.github.com/janavikasera)                      Boston, MA
*Bronze Medal Winner in Kaggle Challenge*                        *October 2022 – November 2022*
- Built a regression model deploying BERT for encoding input data to predict English Language Learner's (ELL) essay score and ranked **159** across approximately **2600** teams
- Fine-tuned the model with different hidden and dense layers and applied **4 optimization strategies**- gradient accumulation, free embedding, dynamic padding, and uniform length batching to increase model accuracy and training speed.

**Music Generator Using LLMs** (www.github.com/janavikasera)                                      Boston, MA
*Adapting Diffusion-LM to Discrete Music Domain*                        *October 2022 – December 2022*
- Adapted Diffusion-LM framework by incorporating two different transformer architectures - **BERT and ELECTRA-BERT** and conducted comparative analysis to improve music quality and optimization power, reduce overfitting, and lower perplexity.
- Worked with over **10,000+ MIDI files** representation of piano sounds and encoded each musical note in the form of (pitch, velocity, duration) and parsed it into text files.

**Stanford University's Open Data Hackathon**                                      San Francisco, CA
*Winner of Data Modeling Event*                        *April 2021*
- Utilized supervised ML models to quantify the relationship between college students' academic performance and their food choices with an aim to recommend food items that would improve students' nutritional index.
- Created a **novel** concentration index based on **5+** predictors by cleaning and pre-processing **126+** rows of dataset in Python.
- Employed AIC and BIC score method in R to fit Multi-Linear Regression model based on **62+** explanatory variables to get the "best-fit" model and expressed the validity of the model by generating diagnostic plots such as scatter-plot matrix, and others.