
EDUCATION

- Northeastern University** - USA Sep 2019 - Dec 2021
MS, Data Science GPA: 3.95/4.0
- **Teaching Assistant:** Deep Learning & NLP - Mentored students & assisted professors in assessments, exams & grading
 - **Research Work (NLP):** Automated an e-learning client's FAANG's interview preparation system - business understanding, data cleaning, building rule-based and machine learning models, evaluating performance
 - **Coursework:** Deep Learning & Neural Networks, Natural Language Processing, Artificial Intelligence, Data Mining (with Machine Learning), Data Management & Database Design, Probability & Statistics, Computation & Visualization
- University of Delhi** - India Jul 2009 - Apr 2011
MS, Mathematics (Hindu College)
- University of Delhi** - India Jul 2006 - Apr 2009
BS, Mathematics (Ramjas College)
- **Coursework:** Probability, Statistics, Linear Algebra, Calculus, Analysis, Optimization, Operations Research

WORK EXPERIENCE

- NLP Engineer Co-op** - Jeevahealth.ai (Medufin LLC), MA, USA Feb 2021 - Present
- **Open Source Project:** DialogFlow conversion process to simplify chatbot building using a one-liner terminal command which provides full support - input & output context setting, parameter saving, lifespan setting, entity tagging
 - Built command line utilities to extract data from online forums like Reddit, Discord & counselling websites
 - Working on building a conversational AI using RNN & deploying using Flask, Gunicorn & Docker on AWS EC2 instance
 - **Technologies:** Python, Tensorflow, HuggingFace, DialogFlow, Git, Gitlab, Jira, Unix, Docker, Node.js, Flask, AWS
- Machine Learning Engineer Intern** - CITAP (SEWA), TX [Github] Jun 2020 - Aug 2020
- Mentored by:** [Parag Kulkarni](#) (SVP SaaS Engineering, Nutanix) & [Ankur Rastogi](#) (SDE, Amazon AWS)
- Built an end-to-end AI-enabled customer support chatbot as a service integrated with slack & deployed on cloud
 - Scraped & processed the Nutanix public documentation to build knowledge base in JSON format
 - Used keywords extraction & LSTM neural network model to predict responses & their related article headings
 - **Technologies:** Python, Flask, Google Cloud Platform, Neural Networks, BeautifulSoup, NLTK, JSON, Git, Github
- SME - Meritnation, Pearson, S Chand**, India Aug 2011 - Aug 2018
- Understood stakeholder's requirements; designed Profit & Loss statements to help in commissioning profitable educational products; performed weekly analysis of per-user data of Ask-Answer forum (queries of 6.5 million users)

PROJECTS

- CNN/Daily Mail, Media Dataset - Text Summarization** Nov 2020 - Dec 2020
- Built a neural attentional sequence-to-sequence LSTM model that can help generate abstractive summaries of text
- CIFAR-100, Image Dataset - Object Recognition** [Github] [Paper] [Towards Data Science] Jun 2020 - Aug 2020
- Trained CNN model using current state-of-the-art EfficientNetB0 with imagenet weights & achieved accuracy of 82% using image augmentation, early stopping, reduce learning rate on plateau & hyperparameter tuning techniques
 - Displayed top predictions by training the model on NVIDIA GPU & 8vCPUs provided by Google Cloud Platform
- Leading US Wholesale Corporation, Customer Dataset - Predict Member Churn** Feb 2020 - Apr 2020
- Processed 10GB industry data to predict member churn using Logistic Regression, k-NN, Naive Bayes, Decision Tree
 - Using model selection - cross-validation technique achieved highest accuracy of 72.2% from Random Forest model
- COVID-19 Monitor** [App] Feb 2020 - Mar 2020
- Developed interactive R Shiny dashboard to monitor geospatial trends of COVID-19 across the globe
 - Fetched data through an automated ETL pipeline connected to John Hopkins COVID-19 data repository on GitHub
- Property Management System** [Github] Sep 2019 - Dec 2019
- Using SQL server developed database (schema, tables, table-level check constraints, computed column, encryption) to store information of rental services & maintenance for residents & staff as a one-stop solution
- Blogging website - Oyewiki** Jun 2016 - Dec 2018
- Developed a DIY tool for authors to write & publish articles, & earn per view (YouTube for readers & writers)
 - Analyzed website data (extracted from MySQL DB using Python scripts) of 10,000 users, 22,000 articles, millions of views to find top read articles, categories, users, fake views tracking & did marketing to increase website traffic

SKILLS SET

- **Language & Database** - Python, SQL, SQL Server, R (Basic), Java (Basic), HTML (Basic)
- **Library** - Pandas, NumPy, Scikit-Learn, Transformers, Matplotlib, spaCy, TensorFlow, Keras, PyTorch, Flask
- **Tool** - Git, GitHub, GitLab, GCP, AWS (Basic), Tableau, JupyterLab, Colab, Docker, Excel, Jira, R Shiny, PyCharm
- **Algorithm** - Regression, Classification, Clustering, Machine Learning, Deep Learning / Neural Networks