Vikash Kumar Prasad

Norman, US 73071 | +1 4059867018 | vikash.k.prasad-1@ou.edu

Summary

Seasoned Data Scientist/current PhD experienced working with large datasets, breaking down information and applying interpretations to complex business concerns. Proficient in NLP, distribution, predictive and hypothetical modeling. Bringing 5.5+ years of related experience strengthening company operations.

Publications:

- PepProEx-Peptide And Protein Exploration Framework, IEEE Vis(Bio Med Vis Challenge), 2023,code

Skills

- Languages: Python2.7, Python3.6, Bash, Js
- OS: Linux, UNIX
- CI/CD Tools: AWS, Jira, Jenkins, Github, AzureML, AzureDataFactory, Docker
- Libraries/Frameworks: Pandas, Numpy, Scipy, Scikit-Learn, Keras, Pytorch, Gensim, BeautifulSoup, Pyspark
- Databases: SQL, MongoDB
- Search-Based Platforms: Elasticsearch
- ML Algorithms: Linear Regression, SVMs, KNN, Logistic Regression, Decision Trees, Random Forest
- Gradient Boosting, K-Means, CRF, LDA

- Deep Learning Algorithms: CNN(Text Classification), LSTM(Text Classification),BiLSTM(Entity
- Recognition), BERT, Bio-BERT
- Statistical Analysis
- Data Mining
- Advanced Mathematical Abilities
- Machine Learning
- Deep Learning
- Time-Series Analysis

Experience

RESEARCH ASSISTANT | 08/2022 – 10/2022
Brookhaven National Lab, US
- Designed end to end ML Tool for training and testing YoloV7
- Designed end to end ML Tool for training and testing REID(Person identification).

DATA SCIENTIST | 12/2020 - 04/2022
Policybazaar Inc - India
- Developed predictive models using machine learning, natural language and statistical analysis methods.
- Developed a Question Answering model to extract entities profession, income, date of birth, education from Audio Speech Recognitions transcripts.
- Built compelling data stories and visualizations to influence decision makers.
Sample Work:
- Train, test and deploy in production NLP based models which process on an average of a 1mn+ sentences daily spoken by agent and customer.
- Developed a Named Entity Recognition model on Hinglish(code switch) text to filter product, diseases, benefits and ordinals.

- Algorithm used: BiLSTM+CRF Framework: Pytorch ii
- Algorithm Used: XGBoost, BiLSTM+CRF

DATA SCIENTIST | 09/2019 - 12/2020
GEP Worldwide - Mumbai, India
- Train, test and deploy in production NLP and big data based models which serve 10-15 high premium customers
- Built and deployed a text classification system that would take in a line description and classify it into the 16000 different classes with accuracy of close to 85%.(Patent, refer publications) Algorithms: (ConvNets)+ Word2Vec Framework: Pytorch ii
- Single-handedly built and deployed automated news gathering based on specific search words around clients and return important and recently published news articles
- Built and deployed for multiple clients including pharma orgs Algorithms:(LDA-Based Topic Modelling, SIF Based Text Summarization)

DATA SCIENTIST | 09/2016 - 08/2019
Gainsight Inc - Hyderabad, India
- Train, test and deploy in production NLU(Sally Gainsight) and churn models.
- Implemented question answering T2T bot Layer which recognized entities and intent from a question on slack.
- Framework: RASA-NLU Algorithm : BiLSTM + CRF.
- Conduct a thorough investigation of data and build churn predictive models for 5-6 different clients.
- Algorithms: Logistic Regression, Random Forest, Gradient Boosting Framework: Scikit-Learn.

DATA SCIENCE INTERN | 09/2015 - 08/2016
Gainsight Inc - Hyderabad, India
- Developed and maintained data pipelines to automate ETL processes.
- Contribute to the building, testing, and deploying predictive machine learning models
- Qualified for Galvanize BootCamp in Data Science

Education and Training

University of Oklahoma - Norman, OK | Ph.D.
Computer Science, Data, Computer Architecture, Expected in 08/2027
- Currently the focus is on explainable machine learning, we are trying to interpret sequence to sequence models for purpose of classification in the case of protein classification(pathogenic/non-pathogenic)
- Completed Course >Data Network, Computer Architecture, Machine Learning, Parallel Distributed Networks
- [4.0] GPA

IIT-ISM - India, India | M.Phil
Applied Mathematics, 05/2014
- MPhil Year Long Recipient
- 8.33 GPA

IIT-ISM - Dhanbad, India | M.Sc
Mathematics & Computing, 05/2013
- Merit-cum-Means Recipient
- 7.69 GPA

Websites, Portfolios, Profiles

- https://vikashkrokc.github.io/
- https://www.linkedin.com/in/vikash-prasad-703771b6/