Vikash Kumar Prasad

1800 Beaumont Dr, Campuslodge 1314 B. •Norman, US 73071 • +1 4059867018 • vikash.k.prasad-1@ou.edu

---

Education

1. Aug,2022-Current, University of Oklahoma, US, Phd Computer Science
    a. 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).
    b. Course Work–GPA: 4.0
        i. Data Networks
        ii. Computer Architecture
        iii. Machine Learning
        iv. Parallel Distributed Networks
    c. Project Work with DoE(BrookHaven National Lab):
        i. Implementing yolov7 to detect objects in real-time.
        ii. Integrating optical flow with yolov7, to spot any moving object changes.
        iii. Implementing ReID to find similar objects in two frames, Integrating with visualization.
    d. Research
        i. Using Transformers to explain classification of protein rna sequence(ongoing), paper submitted at CGA.
2. Aug, 2013-May,2014, IIT-ISM Dhanbad, India, M.Phil Applied Mathematics
3. Jul, 2011-May,2013, IIT-ISM Dhanbad, India, M.Sc Mathematics & Computing
4. July,2007-May,2010, University of Burdwan, India, B.Sc Mathematics(Hons)

Experience

1. Dec, 2020-April, 2022, Policybazaar Inc, Gurugram India, Data Scientist
    a. Train, test and deploy in production NLP based models which process on an average of a 1mn+ sentences daily spoken by agent and customer.

        i. Developed a Named Entity Recognition model on Hinglish(code switch) text to filter product, diseases, benefits and ordinals.

            1. Algorithm used: BiLSTM+CRF

            2. Framework: Pytorch

        ii. Developed a Question Answering model to extract entities profession, income, date of birth, education from Audio Speech Recognitions transcripts.
            1. Algorithm Used: XGBoost, BiLSTM+CRF
2. September, 2019-Dec, 2020, GEP Worldwide, Mumbai India, Data Scientist

a. Train, test and deploy in production NLP and big data based models which serve 10-15 high premium customers.

    i. 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)
        1. Algorithms: (ConvNets)+ Word2Vec
        2. 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
        1. Algorithms:(LDA-Based Topic Modelling, SIF Based Text Summarization)

3. September, 2016-August, 2019, Gainsight Inc, Hyderabad India, Data Scientist
    a. Train, test and deploy in production NLU(Sally Gainsight) and churn models.

        i. Implemented question answering T2T bot Layer which recognized entities and intent from a question on slack.
            1. Framework: RASA-NLU
            2. Algorithm : BiLSTM + CRF
        ii. Conduct a thorough investigation of data and build churn predictive models for 5-6 different clients.
            1. Algorithms: Logistic Regression, Random Forest, Gradient Boosting
            2. Framework: Scikit-Learn
4. September, 2016-August, 2019, Gainsight Inc, Hyderabad India, Data Science Intern
    a. As an intern, I learned the major concepts of Data Science and about how data science is applied in Gainsight.
        i. Contribute to the building, testing, and deploying predictive machine learning models.
        ii. Qualified for Galvanize BootCamp in Data Science(2016-May ,2016-Aug)

Technical Expertise

1. Languages: Python2.7, Python3.6, Bash, Js
2. OS: Linux, UNIX
3. CI/CD Tools: AWS, Jira, Jenkins, Github, AzureML, AzureDataFactory, Docker
4. Libraries/Frameworks: Pandas, Numpy, Scipy, Scikit-Learn, Keras, Pytorch, Gensim, BeautifulSoup, Pyspark
5. Databases: SQL, MongoDB
6. Search-Based Platforms: Elasticsearch
7. ML Algorithms: Linear Regression, SVMs, KNN, Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, K-Means, CRF, LDA
8. Deep Learning Algorithms: CNN(Text Classification), LSTM(Text Classification),BiLSTM(Entity Recognition), BERT, Bio-BERT

Research Interests

1. Explainable Machine Learning
2. Visualization
3. Causal Inference

Publications

1. Title: Data Processing And Classification
    a. Authors: Subhash Makhija, Santosh Bhat, Nishanth Koganti, Venkata Sri Harsha Vemuluru, Vikash Prasad
    b. US Patent No: US20210182659A1.

Social Media

1. Linkedin: https://www.linkedin.com/in/vikash-prasad-703771b6/