Paritosh Katre

[katreparitosh@gmail.com] | [GitHub] | [Personal Page] | [LinkedIn] [Pune, INDIA] | [+918806491899]

EDUCATION

Vishwakarma Institute of Information Technology, Pune University — Pune, India
Bachelor of Engineering in Computer Engineering (GPA: 8.55/10) — Aug '16 - May' 20

SKILLS

Languages: Python | Java | C++, **ML/DL:** Ktrain | TensorFlow | Keras | PyTorch | scikit-learn, **NLP:** NLTK | fast-bert | Stanford CoreNLP | spaCy | Transformers | Gensim **Data Science**: Pandas | NumPy | Matplotlib | Plotly | beautifulsoup4 | Selenium | eli5 | Flask, **Cloud:** AWS EC2 | AWS SageMaker | AWS S3 | **MLOps:** Docker, Gevent, **Annotation:** Label Studio | Doccano, **Databases:** Neo4j | MySQL | SQL | MongoDB

EXPERIENCE

Citispotter Limited *(Remote)* — Cambridge, UK
NLP Lead Engineer — Oct '20 - Present

- Leading the end-to-end research, development and deployment of NLP Systems for Product.
- Designed the real-time ML product pipeline capable of **Continual Learning** and **User Feedback Loops.**
- Performed **multi-GPU training** on **Amazon SageMaker** by connecting **S3 buckets** and docker images from **AWS ECR**.
- Evaluated NLP models for **Bias & Fairness, Behavioral - Adversarial Testing,** and **Linguistic Capabilities** before shipping models to production environments at scale.
- Designed and productionized ML **REST APIs** on **EC2** (CPU/GPU) instances by doing load/performance testing on **POSTMAN, JMeter** and **Locust.**
- Developed NLP algorithms around **POS Tagging, Named Entity Recognition (NER), Gradient-Attributions, Word Embeddings Similarity, WordNet, Phrasal Paraphrasing**.
- Proposed **data augmentation** as a strategy for data creation using tools like using **NLPAug, TextAttack,** and **TextAugment.**
- Performed short/long-term infrastructure budget plans on AWS to reduce effective costs by **20-30%.**
- Reduced real-time inference latency of model serving by **64%** using **CUDA MPS, multi-GPUs,** and **Gevent App Server**.

Citispotter Limited *(Remote)* — Cambridge, UK
NLP Intern — June '20 - Sept' 20

- Researched and trained **15+** SOTA models like BERT, DistilBERT, and CNN for text classification problems in Emotion Detection, Aspect-based Sentiment Analysis, Hate Speech Detection, Gender Neutrality, etc.
- Experimented with **Explainable AI** techniques like **LIME, eli5**, etc.
- Deployed the models wrapped in **Docker, Nginx, Gunicorn** on **AWS EC2** as interactive **Flask Apps**.
- Performed **linguistic resource creation** in terms of datasets, lexicons, research papers.
- Utilized data annotation tools like **Label Studio, Docanno** to experiment with **Active Learning** & **Pre-labelling** pipelines.

C-DAC, Scientific Society under Government of India — Pune, INDIA
Research Intern (NLP) — April '19 – May '20

- Developed a "Pre-crime Tool" that predicts potential criminal activities from English Unstructured Texts using a hybrid Classical NLP-ML approach.
- Implemented **NER, Anaphora, Co-reference Resolution,** and **Dependency Parsing** to extract **Relation Triples** for developing **Knowledge Graph** using **Neo4j** and **networkx.**
- Designed a **Weighted Sum Model (WSM)** for the prediction module.
- Tech Stack: Stanford CoreNLP, NeuralCoref, spaCy, Neo4j, Flask, etc.

RESEARCH PUBLICATIONS

- [Springer] [Paper] - Paritosh D. Katre, et al., *"Survey and Gap Analysis on Event Prediction of English Unstructured Texts"*, Springer SIST - Fourth International Conference on ICT for Intelligent Systems (ICTIS 2020), 15th-16th May 2020, ISSN: 2190-3018, Ahmedabad, India.
- [IEEE] [Paper] [Code] - Paritosh D. Katre, *"Text Mining and Comparative Visual Analytics on Large Collection of Speeches to Trace Socio-Political Issues"*, 2019 IEEE 9th International Advance Computing Conference (IACC 2019), 13th -14th December 2019, Tiruchirappalli, India.
- [Scopus] [Paper] [Code] - Paritosh D. Katre, *"NLP based Text Analytics and Visualization of Political Speeches"*, International Journal for Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-3, September 2019, pp. 8574-8579.
- [Draft Stage] *"Pre-crime Tool: Attack Prediction from Unstructured English Corpora"*.

AWARDS

- **Best Paper Award** and **Cash Prize** at NCRAC -2019, Coimbatore.
- **Lead Paper Award** of the NLP-Session at IEEE IACC 2019, Tiruchirappalli.