

Lead Data Scientist with ~10 years of experience in **Machine Learning, Deep Learning (NLP)** and Software Engineering. Known for not only leading, but also cultivating **high-performance teams**, resulting in consistently remarkable achievements.

EMPLOYMENT & PROJECTS

DRAUP Inc. (Oct 2018- present), **Lead Data Scientist - Architect**

- ❖ **Leading 12 member Data Science team**, guiding at selecting right cutting-edge Machine Learning/Deep Learning technologies..
- ❖ Engaged in **effective collaboration** with diverse cross-functional teams to tackle both internal & client-facing challenges by leveraging ML.
- ❖ Successfully **deployed over 15+ ML/DL models** in production (**AWS ECS**) & leveraged **Apache Spark**, GPUs to address large-scale business challenges (**AWS EMR**).
- ❖ **News Processing Engine**: A system, similar to google news, powered by language models to process company specific news at scale.
 - Trained a **Language Model** (RoBERTa), on 550M tokens (called NEWSoBERTa), achieving a **60% reduction** in perplexity score on Mask Language Modelling (MLM) task and successfully applied to downstream tasks such as **classification and Named Entity Recognition (NER)**, **semantic embedding generation** problems etc..
 - Developed an in-house **scalable clustering algorithm** on the same base model, to eliminate near **data duplication** and solve multiple use cases including **trend recommendation engine**.
- ❖ Trained a **BERT**-based sentence-transformer (SBERT) (combination of **cross-encoder and bi-encoder attention**) model to match any job title to fixed job roles, addressed a critical legacy issue of the company.
- ❖ Trained Seq2Seq BART Encoder-Decoder **Transformer model** to generate paraphrase sentences with **91% accuracy** and eliminated complete manual dependency.
- ❖ Designed a ML system (**XGBoost & Random Forest**) for automated **information retrieval (IR)** from domain-specific URLs, reducing human effort by 75% and engineered multiple ML tools for **effective annotation** of training data, including an [HTML Tag Annotator](#)
- ❖ Designed and implemented a solution utilizing **coreference resolution and customized NER** to effectively tackle pressing customer concerns.

CALIENT Technologies Inc., Software Engineer - **Machine Learning** (Oct 2017 - Sep 2018)

- Built Machine Learning models to categorise network device error logs into multiple categories (**word2vec + logistic regression**) and implemented an alarm system using **Java** for quick response.

COGOPORT *MUMBAI*, Software Developer (Feb 2017 - Sep 2017)

- Designed & Developed Bid Request System including **data analytics pipeline** and API development using **Ruby on Rails**.

TECHNICHE e-commerce solutions, Full Stack Developer (Apr 2015 - Feb 2017)

- Proficiently designed and developed multiple **e-commerce websites** using **Node JS (MEAN stack)**, encompassing essential features such as user authentication, authorization, shopping cart functionality, and comprehensive order management

Hewlett-Packard, Technical Consultant (Aug 2013 - Dec 2014)

- Developed **Java** applications for internal use, utilizing extensive SQL expertise for banking project

Skills, Languages & Technologies

- **Python, Apache Spark, PyTorch, TensorFlow, HuggingFace** 🤗, **Scikit-Learn, Keras, Spacy, BetterTransformer,**
- Traditional ML algorithms & Deep learning algorithms like **LSTMs, Transformers (BERT, BART, RoBERTa, GPT)** etc
- AWS ecosystem: EMR, Sagemaker, ECS, S3, EC2.

EDUCATION

Bachelor of Engineering (B.E.) in Computer Science (2009 - 2013) - VTU