

[sachinkalsi15@gmail.com](mailto:sachinkalsi15@gmail.com),

+91 9945746470

**SACHIN KALSI**, *problem solver*

[YouTube](#) | [Portfolio](#) | [Blog](#) | [LinkedIn](#) | [GitHub](#)

Data Scientist Lead with ~8 years of experience in Machine Learning, Deep Learning (**NLP**) and Software Engineering

## EMPLOYMENT & PROJECTS

---

### **DRAUP Inc.** (Oct 2018- present), Data Scientist IV - Team Lead

- Leading and managing the Data Science team of 12 members. helping the DS team to choose the right ML technology.
- Collaborated with different cross-functional teams to address internal & client-facing problems using ML.
- Successfully **deployed over 15+ ML/DL models** in production (AWS [ECS](#)) & **used Spark to address large-scale business challenges** (AWS [EMR](#)).
- **Fine-tuned Seq2Seq BART Transformer model** to generate paraphrase sentences with **91%** accuracy & eliminated complete manual dependency.
- Developed an in-house **scalable clustering algorithm** to eliminate data duplication and solve multiple use cases including **trend analysis**
- Designed a ML system (**XGBoost & RF**) for automated **information extraction** from domain-specific URLs, reducing human effort by 75%.
- Engineered multiple ML tools for **efficient annotation of training data** for information extraction, including an [HTML Tag Annotator](#)
- Trained a **BERT-based sentence-transformer** model to match any job title to fixed job roles, **addressed a critical legacy issue of the company**.
- Trained **RoBERTa** models on 230M tokens, achieving a **60% loss reduction** in MLM task and successfully applied to **downstream tasks** such as **classification, NER, and Q&A**.
- Designed and implemented a solution utilizing **coreference resolution and customized NER** to effectively tackle pressing customer concerns

### **CALIENT Technologies Inc.**, Software Engineer - ML (Oct 2017 - Sep 2018)

- Built ML models to classify device network error logs into multiple categories (**word2vec + logistic regression**) and implemented an alarm system for quick response, integrated with a Java application.

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

- Contributed to the development of a Bid Request System, data analytics pipeline, and REST APIs using Ruby on Rails.

## **Techniche e-commerce solutions**, Full Stack Developer (Apr 2015 - Feb 2017)

- Designed & implemented various features, including user authentication and authorization, shopping cart management, & order processing, for multiple e-commerce websites using the MEAN stack with end-to-end ownership.

## **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, PySpark, Scikit-Learn, Keras, TensorFlow, Hugging Face 🤖, PyTorch, Spacy
- Traditional ML algorithms & Deep learning algorithms like LSTMs, Transformers (BERT, BART, RoBERTa) etc
- AWS ecosystem: EMR, Sagemaker, ECS, S3, EC2.

## **EDUCATION**

---

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