sachinkalsi15@gmail.com,
+91 9945746470

SACHIN KALSI, *problem solver*

YouTube | Portfolio | Blog | LinkedIn | GitHub

Senior Data Scientist with ~8 years of experience in NLP, Machine Learning, Deep Learning and Software Development

EMPLOYMENT

DRAUP Inc. (Oct 2018- present)

Data Scientist IV - Team Lead (Jan 2022 - present)

- Leading, managing and mentoring a team of 12 data scientists.
- Collaborate with product managers, engineering teams & other teams to design and deliver high quality solutions & helping the DS team to choose the right ML technology

Data Scientist III (Jan 2020 - Dec 2021)

- Trained Seq2Seq Transformer model to generate paraphrase sentences with 91% accuracy.
- Developed a new clustering algorithm to group together similar datapoints (dynamic clustering) to solve multiple use cases

Data Scientist (Oct 2018 - Dec 2019)

- Designed a Machine Learning system to extract relevant and specific information from a given URL, which reduced the human efforts by ~70%

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

- Analyse & build ML models to classify the device network error logs into multiple categories & raise an alarm if required

COGOPORT, Software Developer (Feb 2017 - Sep 2017)

- Designed Bid Request System, data analytics pipeline & involved in development of REST APIs

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

- Developed e-commerce websites using MEAN stack with end to end ownership

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

- Built internal Java applications

PROJECTS

Data Deduplication at Scale

- Developed a scalable dynamic clustering algorithm to group together similar data points to remove the duplication of data points and to pick the relevant datapoint from the cluster
- The model has resulted in reduction of manual effort by 85%

NLP - Generate paraphrasing sentences

- Fine-tuned BART Seq2Seq Transformer model to generate paraphrase sentences without grammar or spelling mistake

NLU Engine

- Designed and developed an NLU Engine that helps the user to navigate the platform
- Various ML models have been trained, developed, tested and deployed for intent classification & named entity recognition (NER)

Information Extraction from Web Pages

- Used XGBoost & Random Forests to extract specific information from a given URL (domain specific) which reduced the human efforts by 67%
- Various ML tools(browser extension, flask app etc) have been developed to increase the efficiency of annotating the training data

Languages & Technologies

- Python, Java, Javascript, Shell Script, Spark
- Scikit-Learn, Keras, TensorFlow, Hugging Face
- Traditional ML algorithms & Deep learning algorithms like LSTMs, Transformers etc

EDUCATION

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