

AAKASH BAKHLE

Pune, MH · aakashbakhle@gmail.com · +91-8007123240 · github.com/aakashb95
linkedin.com/in/aakashb95

EXPERIENCE

Centre for Development of Advanced Computing (C-DAC)

Project Engineer (NLP R&D) - Applied AI Group

Pune

Mar 2021 - Present

- Worked on end-to-end Coreference Resolution on English texts using SpanBERT. Wrote cluster head finding algorithm and resolution logic to handle cataphora and anaphora in input text.
- Modified OntoNotes 5.0 programatically to reduce large clusters while retaining named entities for coreference resolution task. Achieved **77%** using AllenNLP trainer on RoBERTa-large model.
- Generated synthetic data catering to Indian demography, trained [DeepMatcher](#) on said data to identify potential duplicates in real-world data. **Reduced comparisons by 98%** by using phonetic similarity and blocking mechanisms.
- Working on end-to-end 360-degree profiling of person entities from unstructured text. Using BERT based model for relation extraction, building custom dataset for said task.
- Implemented perspective based sentiment analysis using Zero Shot Learning. Explanation can be found [here](#).

Cognizant Technology Solutions

Developer - Oracle Fusion Middleware

Chennai - Bangalore

Nov 2017 - Feb 2020

- Wrote Oracle Data Integrator (ODI) code for data migration from Oracle DB to Microsoft SQL Server and Java Message Service to XML to Oracle DB for a retail client based out of the middle-east.
- Led a team of 3 members, deployed ODI code in production **without errors or escalations from clients**.
- As a part of the performance tuning team for a US based retail client, **reduced running time of two migration jobs** by 70% and 83.33%.

EDUCATION

Centre for Development of Advanced Computing (C-DAC)

PG Diploma Artificial Intelligence - *Grade: A*

Pune

Feb 2020 - Feb 2021

Smt. Kashibai Navale College of Engineering

B.E. Computer Engineering - *72%*

Pune

Aug 2013 - Jul 2017

PROJECTS

Cancer Detection using Gene Expression

- Analyzed and batch processed zip files from The Cancer Genome Atlas (TCGA) database using python scripts to create a csv file containing **60,000 features and 11,000 rows**, where each feature represented a unique gene and each row was a unique patient.
- Designed a voting classifier based on PCA+CNN, Cosine Similarity+XGBoost, UMAP+LGBM to obtain a robust binary classifier which had an **f1-score of 0.98** and **accuracy of 0.97**.

Car make and model detection

- Built an image classification model using FastAi to detect make and model of a car based on Stanford Car dataset consisting of 16000 images of 196 different classes.
- Dockerized the application and deployed it on an Azure compute instance and AWS Fargate.

LANGUAGES AND TECHNOLOGIES

- Python
- Numpy, Pandas, Scikit-learn, Pytorch, SpaCy, HuggingFace
- Git, Flask, FastAPI, Streamlit, Docker