

VINAY KAUNDINYA RONUR PRAKASH

Marienstraße 26, Paderborn 33098

☎ +49 17642728866 ✉ vinaykaundinya95@gmail.com

Education

Universität Paderborn

Master of Science in Computer Science | Majors: Intelligence and Data

Apr. 2020 – Current

Paderborn, Germany

PES Institute of Technology, Bangalore South Campus

Bachelor of Engineering in Information Science and Engineering

Aug. 2013 – July 2017

Bangalore, India

Experience

Universität Paderborn

Werkstudent - Research at Computational Social Science (WHB)

Aug 2021 – Current

Paderborn, Germany

- Worked on "Bias in AI models", which is a joint project by the Paderborn University and Bielefeld University in support of the Joint Artificial Intelligence Institute (JAI).
- Building a framework/pipeline structure for evaluating custom-trained LLMs (Word2Vec, Glove etc.) using social bias metrics (such as WEAT, RNSB etc.). Allows to mix and match different datasets, word embedding models, queries and social bias metrics.
- Experience in handling medium and large-sized text datasets (Pile, Openwebtext etc.), performing tokenization, removing noise, handling missing values and addressing data quality issues.
- Evaluating the impact of concurrent workers on reproducibility, varying embedding sizes, window sizes and min_count of words using social bias metrics.
- Visualizing the produced results using tools (plotly, matplotlib etc.) in order to show the influence of a parameter on a metrics value.

Deloitte Consulting India(Offices of US)

Business Technology Analyst

Jan. 2018 – Mar. 2020

Bangalore, India

- Design and implementation of Data Warehouse migration solutions. Designed data migration scripts, and drove data modelling, analysis and visualization of large data sets.
- Translating customer requirements into efficient RICEFW, CFin objects on SAP ABAP, for global clients in industries like chemical manufacturing, Healthcare etc.
- Working closely with functional teams from SAP functional modules like FI (Finance), SD (Sales and Distribution), TM (Transport Management), MM (Material Management) etc.
- Acted as Point-of-contact between dev teams internationally and Served as Scrum Master for the internal team.

Projects

Evaluating Data-Driven Approaches to Improve Word Lists for Measuring Social Bias in Word Embeddings

| *Master Thesis*

Oct. 2022 - Apr. 2023

- Research on Word embeddings trained on massive amounts of human-produced text that not only capture human semantics but also end up encoding various negative societal biases and stereotypes prevalent in the text.
- Based on Psychology experiments (IAT), evaluated factors of two main Social Bias metrics (WEAT and RNSB) that influence the measured bias in Word Embeddings.
- Developed and evaluated 3 data-driven methods based on classification and clustering techniques for generating Query Word Lists.
- **Technologies** *Scikit learn, Tensorflow/Keras, Gensim, HuggingFace Datasets, Python, Dask, PySpark*

ML4PdM : Machine Learning 4 Predictive Maintenance | *Master Project*

Apr. 2021 - Jan. 2022

- Developed an easy-to-use Python library to tackle maintenance problems like Health Index (HI) Estimation and Remaining Useful Lifetime (RUL) estimation.
- Evaluated pipelines using the Turbofan Engine Degradation Simulation (C-MAPSS) dataset and Prognostics and Health Management (PHM08) datasets provided by NASA Ames Prognostics Center of Excellence (PCoE).
- Set up CI/CD Pipeline including SonarQube and Sphinx Python Documentation Generator.
- **Technologies** *Scikit learn, Tensorflow/Keras, TSFresh, PdMFF, Python*

Podcast InsightBot

August 2023

- An AI chat app that can answer questions by extracting and analysing specific Youtube videos shared as URLs.
- The transcript of the video shared is split into chunks, transformed into embeddings and further ingested into an LLM (GPT-3) along with user prompts to answer a specific question on the ingested data.

- **Technologies:** LangChain (Chains, Agents, Memory, Document Loaders, Text Splitters, Vectorstores etc), OpenAI (GPT-3), FAISS, Streamlit

Beer Recommendation System

April 2022

- Built neighborhood based and model based collaborative filtering algorithms for beer recommendation.
- Developed a content-based recommendation system using NLP on a dataset of 1.5 Million reviews of Beer from BeerAdvocates.
- Tackled recommendation concerns such as cold- start, variety, serendipity and accuracy.
- **Technologies** *Python, Surprise, NLTK, Scikit learn*

Plottr

October 2020

- Used Python tools to extract correlations between different data points in PUBG dataset (consisting of 1000+ matches) and present them using visualization tools.
- Using libraries like Plotly, ggplot, reshape and others to provide a visualization of different aspects that are factors for winning the game in the given dataset.
- Leveraged Streamlit to create a web app for displaying the visualizations.
- **Technologies** *Plotly, ggplot, Scikit learn, Power BI, REST API*

Certifications

Practical Data Science on the AWS Cloud Specialization

Aug. 2023

Deeplearning.ai | AWS | Coursera

- Prepare data, detect statistical data biases, perform feature engineering at scale to train models, and train, evaluate and tune models with AutoML.
- Build, deploy, monitor and operationalize end-to-end machine learning pipelines.
- Store and manage ML features using a feature store and debug, profile, tune, and evaluate models while tracking data lineage and model artefacts.
- Build data labeling and human-in-the-loop pipelines to improve model performance with human intelligence.

Technical Skills

Languages: Python, C, C++, JavaScript, SQL, SAP ABAP, HTML, CSS

Technologies/Frameworks/Libraries: AWS (S3, Glue, Athena, SageMaker), Scikit Learn, Tensorflow/Keras, LangChain, Huggingface, Gensim, AutoML, Numpy, Pandas, Dask, TSFresh, PySpark, Spacy, Power BI, UiPath, Matplotlib, Plotly, Seaborn, NLTK, Linux, Docker, Git (CI/CD)

Relevant Coursework

- | | |
|--|-----------------------------------|
| • Machine Learning | • Model-Based Systems Engineering |
| • Data Science in Industrial Applications | • Software Quality Assurance |
| • Data Science for Physics and Engineering | • Networked Embedded Systems |
| • Computational Argumentation | • Cooperative Mobile Systems |

Extracurricular

Robotic Process Automation Challenge

Sep. 2019

Deloitte Consulting India(offices of US)

- Developed a Bot using UiPath, to automate redundant project management activities in the internal project team.
- Activities which took almost 40 man hours a month were reduced to just 8 mins (per week) of Bot processing.