AVNEET KAUR

Data Engineer, Novo Nordisk

- Ørestad Boluevard 36, 2.3, Copenhagen
- +45 31833156
- avneet14027@iiitd.ac.in
- linkedin.com/in/avneetkaur97/
- github.com/avneet14027

Skills

Programming Languages

Python • Typescript • SQL • Java • C#

Libraries and Frameworks

CDK • Pytorch • Tensorflow • ScikitLearn • Numpy • Pandas • Nipype • PySpark

Tools and Technologies

AWS Services • Github • Visual Studio • Jupyter • RStudio • Hadoop HDFS • Azure DevOps • IBM Bluemix • Docker • Neo4j • Confluence • JIRA • Overleaf • MySQL • PostgreSQL

Technical Coursework

- Large Scale Data Analytics
- Machine Learning
- Data Mining
- Artificial Intelligence
- Database Management Systems
- Signal and Image Processing
- Natural Language Processing
- Probability and Statistics
- Algorithms & Data Structures
- Design and Analysis of Algorithms
- Information Retrieval
- Network Science

Publications

- Multidimensional Analysis of Trust in News Articles, AAAI , Paper Link
- That's Interesting, Tell Me More! Finding Descriptive Support Passages for Knowledge Graph Relationships, **ISWC,** Paper Link
- FlavorDB: a database of flavor molecules, Nucleic Acid Research, Paper Link

Summary

I am currently working as a Data Engineering Professional at Novo Nordisk. My work entails building data pipelines by developing and running ingestion and transformation processes to make data available for analysis. Prior to this, I graduated with an M.Sc degree in Computer Science from the University of Copenhagen, Denmark. I am inspired by the advancing developments in making computers imitate and understand human language, vision, and intelligence and passionate about building tech for social good. My core competencies include a strong programming background acquired through the various courses and internships that have enabled me to gain hands-on experience in machine learning, software development, and version control.

Education

Sept 2019 - January 2022
Masters of Science, Computer Science
University of Copenhagen, Denmark

August 2014 - August 2018
Bachelors of Technology, Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi, India

Work experience

February 2022 - Present
Data Engineering Professional
Novo Nordisk

- Support the onboarding and curation of data in the NNEDL (Data Lake)
- Build data pipelines: develop and run ingestion and transformation processes to make data available for analysis.
- Communicate and collaborate with our business partners in the Digital Health team.
- Design and implement automated processes for data collection.

Sept 2021 - January 2022
M.Sc. Thesis Student
Neurobiology Research Unit, Rigshospitalet

- Building an efficient data pipeline for preprocessing of Positron Emission Tomography (PET) images

June 2021 - August 2021
Summer Technology Analyst (Machine Learning & Natural Language Processing)
BlackRock, Inc.

- Technology Summer intern at the Aladdin Product Group.
- Developed a classifier using Natural language processing for classification of bond reports as being environment friendly.

July 2020 - April 2021
Student Assistant, (Deep Learning & Image Analysis)
StatuManu ICP Aps, Copenhagen, Denmark

- Developed and Tested deep learning model for detecting optic disk and retinal vein pulsations
- Literature Survey, quick prototyping of models.
- Documentation of development activities to comply with Medical Device Regulation.

March 2019 - July 2019
Research Intern, (Network Analysis & Natural Language Processing)
IBM Research, New Delhi, India

- Trust in the selectivity of facts: By analyzing the textual content of the news articles.
- Trust in journalistic assessment: By analyzing journalists' perspective of a news article and tendency to misrepresent facts.
- Trust in topic selectivity: By analyzing the temporal diversity of the issues being covered by various media houses.

This resulted in a publication at the Association of Advancement of Artificial Intelligence, (AAAI) 2020.

January 2019 - July 2019

Research Assistant, (Natural Language Processing)

Indian Institute Of Technology, New Delhi, India

Developed a platform extract important events from news articles to facilitate:

- Generation of news summaries.
- Clustering of similar events reported by different sources.
- Prediction of future events based on previous occurrences

May 2017 - April 2018

Research Intern, (Information Retrieval)

IBM Research, New Delhi, India

- Developed an algorithm for Evidence extraction for Knowledge graph relation triples using probabilistic information retrieval.
- Our work resulted in a publication at International Semantic Web Conference 2018.
- Won the Best Paper Award (Link)
- Paper Link

Notable Projects

Feb 2021 - June 2021

Delineating field boundaries using sentinel-2 imagery (Project in Practice) (Remote Sensing & Computer Vision)

Work carried out in collaboration with DHI Gras

- Implementing a deep learning model to delineate field boundaries using convolutional networks.
- The idea was to reduce human intervention in agricultural records which may be prone to errors and labor intensive.

Sept 2020 - Jan 2021

Imputation of single-cell gene expression using protein-protein networks (Bioinformatics, Data Analysis)

Project Outside Course Scope, UCPH

- Leveraged structural information from the protein-protein network interaction to impute the drop out values in scRNAseq data.
- Validation and benchmarking of the imputation model developed.

Jan 2018 - April 2018

Compilation of flavor molecules and their taste-based classification (Bioinformatics, Data Analysis)

Bachelors Thesis Project, IIITD

- The goal was to perform taste based classification by analyzing the performance of various Machine Learning algorithms, to understand an aspect of flavor, and also and analyze the taste and flavor profiles of various compounds based on physico-chemical properties.

Sept 2017 - Dec 2017

Extracting Relevant Abstracts in the Biomedical Domain from PubMed (Bioinformatics, Natural Language Processing)

Bachelors Thesis Project, IIITD

The aim was to implement and evaluate various retrieval and ranking methods for the purpose of finding relevant abstracts of scientific articles from PubMed, a database of references and abstracts on life sciences and biomedical topics, compare the effectiveness of such ranking methods.