Davide **Abbattista**

Bitonto, 70032, BA, Italy

✉ davide27abbattista@gmail.com | 🏠 davide-abbattista.github.io | ⬛ davide-abbattista | 💼 davide-abbattista | 🎓 Davide Abbattista

Experience

Data and AI Researcher *Engineering Group, Bari, Italy*

Python, Hugging Face, Git, Research, NLP, LLM, Artificial Intelligence, Data Science *Dec. 2024 - Actual*

- Engage in Research & Innovation projects in diverse national and international contexts, driving solutions that address real-world challenges.
- Study, analyze and apply cutting-edge methods, methodologies, techniques and tools to meet project requirements and objectives.
- Design, develop, and test prototypes leveraging Machine Learning (ML), Advanced Analytics, and Large Language Models (LLMs) to deliver impactful, scalable solutions.

Research Intern *SisInfLab, Politecnico di Bari, Italy*

Python, PyTorch, Hugging Face, Git, Recommender Systems, State Space Models, Academic Research *Jul. 2024 - Nov. 2024*

- Continued the work previously done as visiting researcher at the University of Glasgow to carry out the master's thesis.
- Designed and implemented S6Rec, a structured state space model (SSM) based on Mamba for sequential recommendation.

Visiting Researcher *IR Group, University of Glasgow, UK*

Python, PyTorch, Hugging Face, Git, Recommender Systems, Transformer-based Models, Academic Research *Mar. 2024 - Jun. 2024*

- Delved into the world of recommender systems with a particular focus on sequential recommendations in the music domain.
- Augmented Transformer-based models with personalized popularity awareness, yielding improvements which range from 25.2% to 69.8%.
- Submitted a paper accepted for publication in the RecSys 2024 LBR track.

Education

M.Sc. in Computer Science Engineering *Politecnico di Bari, Italy*

Final grade: 110/100, cum laude *Sep. 2022 - Dec. 2024*

- Curriculum: Artificial Intelligence and Data Science.
- Main Courses: Fundamentals of Machine Learning, Agent-Based Artificial Intelligence, Mathematical and Statistical Methods for Artificial Intelligence, Deep Learning, Information Retrieval and Personalization, Information System Design and Big Data.
- Thesis: Enhancing Sequential Music Recommendation with Personalized Popularity Awareness.

B.Sc. in Computer Science and Automation Engineering *Politecnico di Bari, Italy*

Final Grade: 110/110, cum laude *Sep. 2019 - Jul. 2022*

- Curriculum: Computer Systems and Applications.
- Main Courses: Algorithms and Data Structures in Java, Mathematical Analysis, Computer Science for Engineering, Databases and Information Systems, Numerical Calculation, Operating Systems, Software Engineering and Web Foundations.
- Thesis: Analysis and modeling of a blockchain-based platform for vehicle management.

Publications

Enhancing Sequential Music Recommendation with Personalized Popularity Awareness

18th ACM Conference on Recommender Systems (RecSys '24), October 14–18, 2024, Bari, Italy

- This paper introduces a novel approach that incorporates personalized popularity information into sequential recommendation. By combining user-item popularity scores with model-generated scores, the proposed method effectively balances the exploration of new music with the satisfaction of user preferences.

DIVAN: Deep-Interest Virality-Aware Network to Exploit Temporal Dynamics in News Recommendation

ACM RecSys Challenge 2024 (RecSys Challenge '24), October 14–18, 2024, Bari, Italy

- This paper introduces DIVAN (Deep-Interest Virality-Aware Network), the proposed solution for the RecSys 2024 Challenge, combining a Deep Interest Network (DIN) for personalized user interest representation with a Virality-aware Click Predictor that utilizes temporal features to estimate click probability based on news popularity.

(Selected) Projects

SciQA - a Scientific Question Answering System with Citations

Python, Hugging Face, LLM, NLP, RAG, Prompt Engineering *Dec. 2024*

- Designed and implemented an end-to-end solution to provide accurate, contextually relevant, and citation-supported answers to user queries.

DistilBERT-based Architecture for Sentiment Analysis

Python, PyTorch, Hugging Face, Matplotlib, pandas *Nov. 2024*

- Designed and implemented a binary sentiment classification architecture for hotel reviews: it includes the pre-trained DistilBERT base uncased model, a dropout layer and a fully connected layer, fine-tuned using a large dataset from Tripadvisor, in order to accomplish the task.
- Obtained an accuracy score of 98%.

Shipping Delays Analysis and Prediction

PySpark, SQL, pandas, Power BI, Data Science, Data Cleaning, Data Analysis, Feature Engineering *Oct. 2024*

- Performed exploratory data analysis to understand the factors influencing shipping delays and identified clusters that bring the highest revenues to the company.
- Developed predictive models (Linear Regression and XGBoost) to forecast shipping delays and quantify the expected delay in days.
- Tailored the training and evaluation process by giving more weight to the most profitable customers.

word2owl – Mapping Natural Language to OWL and Back

Python, ollama, Git, Protégé, LLM, RAG, Prompt Engineering *Aug. 2024 - Sep. 2024*

- Designed and implemented a system, based on LLM and RAG, able to convert natural language sentences into OWL 2 declarations and axioms (insertion mode) and to answer queries in natural language concerning its memorized knowledge (query mode).
- Designed a RAG technique, based on a knowledge graph and a vector index (faiss with embedding based on a SBERT model), to include in the system prompt the available knowledge related to the user's query.
- Designed 4 different prompt strategies for the system's insertion mode: few-shot, few-shot Chain-of-Thought, prompt chaining few-shot, prompt chaining few-shot Chain-of-Thought.
- Defined evaluation measures and created an ontology to manually test the system, both in the insertion mode and in the query mode.

spamBERT – Spam Text Classification Using BERT

Python, PyTorch, Hugging Face, Scikit-learn, pandas, Git, NLP, Fine-Tuning *Jan. 2024*

- Designed and implemented a spam classification architecture for email and sms texts: it includes the pre-trained BERT base cased model (with 12 encoders) and a fully connected layer, fine-tuned using SMS Spam Collection and Spam-Ham Dataset, in order to accomplish the task.
- Implemented a simple user-friendly webpage for model inference.
- Obtained an accuracy score of 99%.

outGANfit – Compatible Outfit Generation Using GANs

Python, PyTorch, NumPy, Matplotlib, Git, GAN, GenAI *Aug. 2023 - Sep. 2023*

- Designed and implemented a Conditional Deep Convolutional Generative Adversarial Networks (cDCGANs-)based architecture, trained with images from the Polyvore Outfits dataset, to generate outfit items compatible with the input garment: it takes as input a t-shirt and generates a matching pant, pair of shoes and accessory.
- Included 3 different GANs, one for each garment to be predicted, composed by 1 generator and 2 discriminators: the 1st used to evaluate the compatibility between the conditioning image and the generated garment; the 2nd used to classify what is generated as real or fake.
- Obtained a FID value of 175.42.

Awards

Scholarship for thesis abroad *Politecnico di Bari, Italy*

BASED ON THE WEIGHTED AVERAGE OF CAREER EXAMS AND THE QUALITATIVE CHARACTERISTICS OF THESIS PROGRAMME *Mar. 2023*

Study award for deserving high school students *High School "Galileo Galilei", Bitonto, Italy*

BASED ON THE AVERAGE VOTES, THE GRADE IN THE STATE EXAMS (100/100 CUM LAUDE) AND THE STUDENT'S PERSONALITY *Dec. 2020*

Study award for 1st place in the local competition of "Olimpiadi di Fisica" *High School "Galileo Galilei", Bitonto, Italy*

ORGANIZED BY THE ITALIAN ASSOCIATION FOR THE TEACHING OF PHYSICS (AIF) *Dec. 2018*

Skills

Languages	Italian (native), English (proficient).
Programming	Python, Java, Prolog, Matlab.
Python Libraries	PyTorch, Hugging Face Transformers, NumPy, Pandas, Scikit-learn, Matplotlib.
Web Development	HTML, CSS, JavaScript, React, Node.js, Express.
Database Management	SQL, MySQL, MongoDB.
Large-Scale Data Analytics	Apache Spark, PySpark, MLlib.
Data Visualization and Analysis	Power BI.