ANDREA BOSIA

andrea.bosia@hotmail.it | +61 492885342 | https://andreabosia.github.io/personalpage/ | 13 Kingston Rd. Sydney, NSW

EDUCATION

University of Sydney *Feb. 2023 -June 2024 (expected)*

Master of Science in Data Science (Major in Machine Learning)

- Relevant Courses: *Advanced Machine Learning, Deep Learning, Database Management Systems, Game Theory, Online Learning, Computer Security.*

Politecnico di Milano *Sept. 2017-Mar 2021*

Bachelor Degree in Computer Science and Engineering

- Relevant Courses: *Data Base, Software Engineering,Technologies for the Web, Data Structures and Algorithms, Fundamentals of Automatic Control.*

WORKING EXPERIENCE

OreFox (Sydney) *Feb. 2024- Present*

Data Scientist

- In the scope of my thesis project I am working with OreFox, a metal mining company, on the deployment of a stand alone news website.
- The final product will aggregate and summarise news articles scraped from the web leveraging a LLM. I am implementing a RAG system via LangChain that queries a vector base.

Brain and Mind Center (University of Sydney) *Dec. 2023- Feb. 2024*

Research Internship

- I worked, together with a team of researchers, to advance results in the field of reconstructing Natural Images from Brain Activities. I worked with the Natural Scene Dataset, a 7T dataset hosted on AWS accessible via S3. I leveraged the COCO dataset to extract image captions.
- I firstly reproduced the results obtained in previous research papers leveraging a node cluster for computational power. I implemented and fine tuned a stable diffusion model for image generation getting fMRI brain signal as input, and conditioned by a textual prompt.

PROJECTS

Class-Conditional Label Noise Robustness 2023

- Working in a team I designed and implemented a classifier robust to class-conditional label noise. I evaluated the performance of two distinguished CNN architectures, ResNet18 and GoogLeNet, when trained with noisy labels.
- I implemented and compared three algorithmic strategies that leverage transition matrices. I investigated the estimation of transition matrices through anchor points.
- I evaluated the performances considering Accuracy, Precision, Recall and F1 score (achieving values ~ 0.9)

Ensemble Model for Multi Label Classification 2022

- In a team of 3 people I developed an ensemble model for multi label classification using PyTorch.
- The dataset consists of images annotated with a textual description. I leveraged the bi-modal input developing 2 classification models, one for the images (a CNN based on ResNet-34) and one for the text (a LSTM).
- After a series of ablation studies and hyperparameters tuning the model achieved an F1-score of 86% finishing in the top 10 of the Kaggle competition.

Desktop Application 2020

- In a team of 3 people I developed a desktop application reproducing the board game “Santorini”. Throughout the project I implemented a multi users, distributed software application provided of CLI and GUI (Java). I adopted a variety of design patterns in order to produce a clean, encapsulated code. I developed functional tests and load tests to validate the final software.

Web Application 2020

- With a teammate I design and implemented an Integrated Retail Management System for an E-commerce. I adopted a Model-View-Controller framework leveraging Java Servlets. I used Thymeleaf in order to provide Dynamic Content Generation. I worked on the DBMS ensuring the efficiency of the relational Data Base and its robustness from malicious attacks.

TECHNICAL SKILLS & TOOLS

Programming Languages: C, Java, JavaScript, Python, R, SQL, HTML, CSS, XML, JSP, Thymeleaf.
Software development: GitHub, Valgrind dynamic analysis tools, GDB debugger, TOMCAT, MySQL.
Big Data & Machine Learning: Numpy, Pandas, Sickit-Learn, Keras, PyTorch, TensorFlow, Seaborn, Tableau, Selenium, AWS EC2
Certifications: Coursera Unsupervised Learning, Recommenders, Reinforcement Learning (Mar. 2023)