ANDREA BOSIA

andrea.bosia@hotmail.it | +61 492885342 | https://andreabosia.github.io/personalpage/ | Bergamo via Bernareggi 4, Italy

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 master thesis project I am working with OreFox, a metal mining company, on the development of of a LLM based news aggregation tool.
- The GenAI tool aggregates and summarises news articles scraped from the web and visualise them in a user friendly News Website.
- The model pipeline consists in a Retrieval Augmented Generation system that queries a vector base.

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

Research Internship

- I collaborated with a team of researchers to advance results in the field of reconstructing Natural Images from Brain Activities.
- I leveraged the Natural Scene Dataset, a 7T dataset hosted on AWS accessible via S3.
- I implemented and fine tuned a text-conditioned latent diffusion model, for image generation using fMRI brain signal as input.
- I showcased the results during a poster session that welcomed both academic and industry attendees.

PROJECTS

Class-Conditional Label Noise Robustness 2023

- I designed and implemented an image classifier robust to class-conditional label noise.
- I tested two distinguished CNN architectures, ResNet18 and GoogLeNet, in combination with various transition matrix based algorithms to train unbiased estimators of the data distribution.
- I evaluated the performances considering Accuracy, Precision, Recall and F1 score (achieving values ~ 0.9)

Ensemble Model for Multi Label Classification 2022

- I developed a multi-modal ensemble model for multi label image classification using PyTorch.
- Leveraged the bi-modal input (images and annotations) I implemented 2 classification models, one for the images (a CNN based on ResNet-34) and one for the text (a LSTM).
- Following series of ablation studies and hyper-parameters tuning the model achieved an F1-score of 86% finishing in the top 10 of the Kaggle competition.

Desktop Application 2020

- I developed a desktop application reproducing the board game “Santorini”.
- I implemented a Java multi users, distributed desktop application provided of CLI and GUI.
- I adopted a variety of design patterns producing clean, encapsulated code. I developed functional tests and load tests to validate the final software.

Web Application 2020

- 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 optimised the Data Base to ensure high throughput and robustness from malicious attacks.

TECHNICAL SKILLS & TOOLS

Programming Languages: C, Java, JavaScript, Python, R, SQL, HTML, CSS, SQL.
Skills & Tools: Git Version Control, Time Series, Supervised&Unsupervised methods, GenAI.
Big Data & Machine Learning: Numpy, Pandas, Sickit-Learn, Keras, PyTorch, TensorFlow, Seaborn, Tableau, Selenium, AWS Cloud Platform (EC2, S3, Amplify).
Certifications: Coursera Unsupervised Learning, Recommenders, Reinforcement Learning (Mar. 2023)