Silvio Pavanetto
silviopavanetto.com
Email : silvio.pavanetto@gmail.com
Github : github.com/reidemeister94

Experience

- **Jobtome** — Stabio, Switzerland
  Software Engineer — *Nov 2021 - Present*
  - **Job Categorizer**: Classification model developed and deployed on **GCP**, using **NLP deep learning techniques (BERT, CNN)**, that processes approximately 1 million of jobs per day.
  - **Job Offers Crawler**: Software architecture written in **Python**, composed by more than **400 crawlers** for many different types of web pages, scheduled and executed in production using **Docker, Airflow and Kubernetes**.
- **Polytechnic University of Milan** — Milan, Italy
  Research Fellow — *Oct 2019 - Oct 2021*
  - **PRIN - HOPE**: HOPE - High quality Open data Publishing and Enrichment - is a PRIN - Progetti di Ricerca di Interesse Nazionale - project financed by the Minister of Education, University and Research, with objective of discovering and exploiting new ways to connect known sources of data with open data.
  - **Periscope**: Map and analyse the unintended impacts of the COVID-19 outbreak by developing solutions and guidance for policymakers and health authorities on how to mitigate the impact of the outbreak.
  - **Italian Museums Reputations**: Data pipeline for ingesting, analyzing and producing insights about online reviews and social media posts of Italian museums.
- **Deloitte Digital** — Milan, Italy
  Software Developer — *Feb. 2019 - Aug. 2019*
  - **Salesforce Projects**: Developer role on international cloud projects using **Salesforce** technology, which includes several programming languages such as **Java** and **Javascript**.

Education

- **Polytechnic University of Milan** — Milan, Italy
  M.Sc. in Computer Science and Engineering; GPA: 3.85 (28.8/30); Final grade: 110L/110 — *Sep. 2016 – Apr. 2019*
- **Polytechnic University of Milan** — Milan, Italy
  Bachelor Engineering of Computing Systems; Final grade: 97/110 — *Sep. 2013 – Sep. 2016*

Selected Publications

- **ICWSM 2021 - Online**: *A Content-based Approach for the Analysis and Classification of Vaccine-related Stances on Twitter: the Italian Scenario;* Collection and analysis of 3 millions of Italian conversations about COVID-19 vaccines on Twitter, investigating the geographical, temporal and lexical distribution of data. Trained a binary classifier that predicts the stance of tweets towards vaccines, i.e., it applies a "Pro-vax" or "No-vax" label. **Link**
- **ICWE 2020 - Helsinki, Finland**: *Generation of Realistic Navigation Paths for Web Site Testing using Recurrent Neural Networks and Generative Adversarial Neural Networks.* Method for generating high-quality weblog data using deep learning techniques, involving comparison of results with a suite of data mining algorithms as a baseline. Main types of algorithms used: Recurrent Neural Networks and Generative Adversarial Neural Networks with a reinforcement learning approach. **Link**

Programming Skills

- **Languages**: Python, Java, Javascript, C++ **Technologies**: GCP, Docker, Kubernetes, Airflow, Pandas, Numpy, Matplotlib, Scikit-Learn, Tensorflow, Keras, Flask, FastAPI.

Projects

- **IDiOM**: Data collection and analysis pipeline for studying COVID-19 related contents and web dashboard available for citizens. **NLP** techniques and relative Python implementations (**NLTK, BERT, word2vec**) involved in the project. Tools used for building microservices architecture: **Docker**. **github link**
- **Vaccinitaly**: Monitor Italian conversations around vaccines on social media (Twitter, Facebook) and understand the interplay between online public discourse and vaccine hesitancy/uptake rates. Techniques and tools used: **MongoDB, Python, Scikit-Learn, Pandas, Transformers, Keras, BERT**. **github link**
- **FantaNBA Predictor**: System for predicting fantasy scores of NBA players in real games, using data science and machine learning techniques. Main languages and frameworks used: **Python, Numpy, Pandas, Scikit-Learn, BeautifulSoap, Keras**.**github link**