Margherita
Gambini

DATE OF BIRTH:

CONTACT




WORK EXPERIENCE

03/2020 - CURRENT - Pisa, Italy

Software developer (part-time)

WEVO s.r.l

- Data gathering and analysis from social media by using Selenium framework.

03/2020 - 06/2020

Master thesis: Developing and Experimenting Approaches for DeepFake Text Detection on Social Media

CNR - IIT Department

- The master thesis is about developing and experimenting approaches for DeepFake text detection on Social Media (Twitter in particular).
- Current language model generators and machine-generated text detectors have been studied.
- I've developed and experimented on a GPT-2 based AI-generated text detector, taking inspiration and tools from Open AI GPT paper, HuggingFace Transformer Library, Simple Transformer Library and Google Colab.

Pisa, Italy

11/2019 - 12/2019

Exploring Movie Recommendation Systems Techniques

Master Degree Project

- Explored the Content-Base and the Matrix Factorization approaches.
- Used Python sklearn library (for evaluation metrics), the matplotlib (for plots), pandas, numpy and pickle libraries (to process data), Surprise library (to test the matrix factorization SVD approach)

05/2019 - 06/2019

Face Clustering

Master Degree Group Project

Development of a tool able to cluster faces based on their deep features' similarity:

- acquired Harry Potter actors' images by using Selenium
- faces extracted from images with a DNN Face Detector available in OpenCV
- the face features were extracted from a VGGFace2 CNN's layer
- reducing the number of features with Python sklearn PCA
- used Python sklearn library for clustering methods (K-MEANS, DBSCAN, HDBSCAN)
- used Tensorflow visualization tool to show clusters applying t-SNE

31/07/2020

Building a Content-Based Image Retrieval index on top of Elasticsearch that can be used to perform a similarity search.

05/2019 - 06/2019

- **User's Authentication Using Information Collected by Smart-Shoes**

  Master Degree Group Project

  This research was about authenticating a person by using just his/her walking-at-normal-pace data. A walkthrough:

  - Gain data collected from 10 volunteers by using two Shimmer3 IMU devices and two devices from the DecaWave TREK1000 kit
  - Used MATLAB to process the gain signals
  - Used the Python Sklearn library to select the features
  - Used the dd_tool MATLAB library to evaluate several one-class classifiers (used Balanced Accuracy, FNR, FPR as evaluation metrics)

07/2018 - 07/2018

- **Development of an ANN and a Fuzzy Inference System to probe the colour difference between a master**

  Master Degree Group Project

  - Used MATLAB Neural Network Tool to devise an Artificial Neural Network able to calculate a difference score between a master colour and a copy colour.
  - Used MATLAB Fuzzy Logic Designer to design a Fuzzy Inference System to output a colour difference score which better resembles the colour human perception.

03/2017 - 04/2017

- **Internship**

  CNR - IIT Department

  - Internship for my bachelor's thesis about the development of a Kibana plugin for choropleth map querying Elasticsearch to retrieve the geo-localized data.

  https://www.iit.cnr.it/ | Pisa, Italy

EDUCATION AND TRAINING

10/2017 - 20/07/2020 - Pisa, Italy

- **Master Degree in Computer Engineering (110/110 cum laude)**

  University of Pisa - Department of Information Engineering

  Thesis title: Developing and experimenting approaches for DeepFake text detection on Social Media

  http://www.ing.unipi.it/it/

31/07/2020

10/2013 - 23/06/2017 – Pisa, Italy

- **Bachelor Degree in Computer Engineering (110/110 cum laude)**

  University of Pisa - Department of Information Engineering

  Thesis title: "Development of a Kibana plugin for choropleth map"

  http://www.ing.unipi.it/it/

09/2008 - 07/2013

- **High School Diploma in Scientific Studies (97/100)**

  Liceo Scientifico "U.Dini"

  https://www.liceodini.it/

LANGUAGE SKILLS

MOTHER TONGUE(S): Italian

English

Listening	Reading	Spoken production	Spoken interaction	Writing
B2	**B2**	**B1**	**B1**	**B2**

DIGITAL SKILLS

Programming and scripting languages

Python 3 / Java 8 / Javascript / SQL / MATLAB

Frameworks

Selenium / Java Server Faces / Primefaces

IDE and tools

PyCharm / Eclipse / Netbeans / Weka / GitHub

DBMS

MySQL / Elasticsearch

31/07/2020

PUBLICATIONS

TweepFake: about Detecting Deepfake Tweets

2020
Collection of the first dataset of *real* deepfake social media messages - which are *real* in the sense that they were actually posted on a social media (Twitter in particular) - to support the deepfake text detection in a real social media setting, where the text generator is unknown and the content is often short. Then, 13 deepfake text detection approaches were tested against the collected dataset.

Developing and experimenting approaches for DeepFake text detection on Social Media (Master Thesis)

2020
Developed a GPT-2 based deepfake text detector and tested it over the dataset of deepfake tweets collected by Fagni et al., 2020 ('TweepFake: about Detecting Deepfake Tweets')

User's Authentication Using Information Collected by Smart-Shoes

2019 https://link.springer.com/chapter/10.1007/978-3-030-34833-5_21
Authentication of a person by using just his/her walking-at-normal-pace data gathered by smart shoes.

LINKS TO PUBLICATIONS

Shared Link to Publications

https://drive.google.com/drive/folders/1yT9rQFj6Qq7l19u7hADfcyJorE0pq48w?usp=sharing

31/07/2020

ACADEMIC COURSES

03/2019 - 06/2019

Multimedia Information Management

- **Text and Audio Retrieval:** Ranked Retrieval, TF-IDF score, Evaluation in Information Retrieval, Audio Retrieval and Fingerprinting,
- **Multimedia Content Representation and Analysis:** Global Features, Local Invariant Feature Detectors, Local Features, Local Features Matching and Aggregation, Face Detection, Deep Learning basis, CNN and recent architectures, Transfer Learning, Regularization Techniques, Representation and Learning, Generative Models,
- **Similarity Indexing and Searching:** Foundation of Metric Space Searching, Exact Similarity Search Methods, The M-Tree Family, Hash-Based Methods, Approximate Similarity Search with M-Tree, Permutation Based Methods, Local Sensitive Hashing
- **Tools**: Elasticsearch, Open CV Java library

10/2018 - 12/2018

Data Mining

- Data Pre-Processing, Association Rules, Sequential Pattern Mining, Classification, Clustering, High Dimensional Clustering, Outlier Detection, Graph Clustering, Constrained Cluster Analysis, Hadoop
- **Tools**: Weka, Python

03/2018 - 06/2018

Intelligent Systems

- The basis of Artificial Neural Networks, Fuzzy Logic and Fuzzy Systems, Genetic Algorithms
- **Tools:** MATLAB Neural Network Tool, Neural Network Clustering Tool, Fuzzy Logic Designer.

31/07/2020