

Niccolò SALVINI

Data Scientist and Engineer | PhD student in Big Data for Healthcare Systems



FORMATION

- 2021 - current** Doctor of Philosophy (PhD) in Big Data for Healthcare System and Research, CATHOLIC UNIVERSITY OF SACRED HEARTS, CAMPUS ROME (MEDICINE DEPARTEMENT), Rome (IT)
- Statistical modelling
 - Machine learning
 - Spatial statistics
 - Empirical health research
- Python R SQL
- 2018 - 2020** Master of science in Statistical and Actuarial Sciences | Data analytics for business and economics, CATHOLIC UNIVERSITY OF SACRED HEART, Milano (IT)
- > 107/110
 - > Thesis : REST Scraping API for Real Estate data, a Spatial Bayesian modeling perspective with INL
 - > Thesis advisor : Prof. Marco della Vedova (Computer Science)
 - Statistical modelling
 - Data management and data mining
 - Statistical Learning and Machine learning
 - Deep Learning and Neural Networks
- Python R SQL Cloud GCP AWS
- 2014 - 2017** Bachelor in Business Economics | Economia e Commercio, UNIVERSITY OF FLORENCE, Firenze (IT)
- > Thesis : IT Products' Price spreads in major online retailers and margins of Arbitrage and Speculation"
 - > Thesis advisor : Prof. Giampiero Maria Gallo (Econometrics)
 - Statistical modelling
 - Micro/macroeconomics
 - Corporate finance
- R SQL GRETL

TEACHING

- 2023** Teaching Assistant, CATHOLIC UNIVERSITY OF SACRED HEART, Rome (IT)
- > Course : Statistics and Big Data
 - > Degree : management of healthcare systems (Magistrale)
- R
- 2021** Teaching Assistant, SPATIAL ECONOMETRICS ASSOCIATION (SEI), Rome (IT)
- > Course : Reproducibility in spatial research (GIT + GitHub)
 - > Spatial Econometrics Advanced Institute (SEI)
- GIT

EXPERIENCE

- In progress** Data Scientist and Engineer Freelance, , Italy / Remote
- 2020** I have been implementing for several small and medium companies in a wide range of sectors :
- > ETL, ELT data Pipelines and Database Management Systems
 - > Machine Learning models and their Production Infrastructures
 - > Webscraping softwares to extract and load data from non conventional sources
 - > Custom automated business reports in various formats
 - > Custom Web Apps explicitly built on top of clients' expectations
 - > Advanced and interactive visualizations
- Python R Docker GCP AWS HTML CSS Javascript Shiny Streamlit React Figma DataWrapper

CONFERENCE AND SPEECH

- > Spatial Econometrics Association conference, Tokyo (JP), remote 2021
- > Spatial Econometrics Association conference, Warsaw (PL), remote 2022
- > "Smart Statistics for Smart Applications" - Conference of the Italian Statistical Society, Milano (IT) 2019
- > "Survey Methods for Statistical Data Integration and New Data Sources" - ITCOSM2022, Perugia (IT) 2022
- > "Spatial Health Econometrics Workshop" - VERM 2022, Venice (IT) Jul 2022

ADVANCED COURSES AND SEMINARS

- > Rafforzamento dell'infrastruttura tecnologica e degli strumenti per la raccolta, l'elaborazione, l'analisi dei dati e la simulazione Dr.ssa Serena Battilomo, Prof. Giuseppe Orbia
- > (CdP) Innovazione digitale e sistemi informativi per il governo delle organizzazioni sanitarie.

LICENSE AND CERTIFICATIONS

- > GMAT Graduate Management Admission Council™ (GMAT™) Release date : gen 2018 · No expiration date 580
- > IELTS (International English Language Testing System) - overall 7.0 Release date : feb 2018 · expiration date : feb 2020 7.0
- > DELF (Diplôme d'études en langue française Release date : apr 2015 · no expiration date 99

RESEARCH PROJECTS

CO.R.E.- CORRUPTION RISK INDICATORS IN EMERGENCY | UE COMMISSION FUNDS AS PART OF THE INTERNAL SECURITY FUND POLICE (ISF-P) PROGRAM WEBSITE

Perugia (IT), 2022

The project under Prof. Gnaldi (PI) was funded for a total of 514 thousand euros. The working group is made up of colleagues Prof. Enrico Carloni, Prof. Marco Mazzoni, Prof. Benedetto Ponti, Prof. Maria Giovanna Ranalli. The international network involves Universitat Oberta Catalunya (Spain), Dublin City University (Ireland), Oficina Antifrau de Catalunya (Spain), Infonodes (Italy), Transparency International (Portugal), Villa Montesca Foundation (Italy). The CO.R.E. focuses on assessing the risk of corruption in public procurement from a preventive perspective. In view of achieving this goal, central to the European agenda, CO.R.E. therefore intends to develop and validate a procedure for the construction of a synthetic indicator of the risk of corruption in an emergency period that can be useful for use by national anti-corruption agencies, the media and citizens for accountability purposes. My responsibilities were to engineer data to scrape and enrich existing databases to serve BI tools and analysis.

Python GCP web scraping

WSF ACADEMY | WALL STREET FOOTBALL WEBSITE

Rome (IT), 2022

Wall street football is an italian startup which leverage first class machine learning and AI to provide the most accurate football players performance predictions. I was selected as a fellow of the Wall Street Football Academy to develop and redefining statistical methodologies applied to sports analytics. In the research scope I developed, together with Dr. Vincenzo Nardelli a data enrichment pipeline for their databases on top of which we applied Deep Learning technique to fuzzy match strings across data sources (mapping each player names across internal databases)

Python R Cloud AWS

INDICATORE DI INFLUENZA SOCIAL | ITEMS WEBSITE

Roma (IT), 2022

The research group has set itself the purpose of analyzing fake news, their impact on vaccination campaigns and the formation of common opinion. This same theme, under the incentive of the director of Items Prof. Cicchetti was involved in monitoring expenditure for advertising campaigns on COVID and similar issues on a regional basis by identified institutional interlocutors in the health sector. The monitoring present on the one hand measures how much has been invested in the correct institutional information in the health sector on the socials of the Meta group (Facebook, Instagram and Whatsapp), on the other hand it serves as an over-alarm bell by measuring the under-representation (or over-representation) of the same phenomenon with respect to the perception on social networks. In this ultimate aim, the hope is to identify an institutional social thermometer with respect to the covid theme by intercepting divergences in communication, trends, and effectiveness.

R GitHub-actions

INDICATORE DI ZONA GIALLA | IITEMS[WEBSITE](#)

Roma (IT), 2021

The research focuses on developing an indicator, namely Indicatore di Stress, which measures the overall stress undertaken by the SSN as a combinatory effect of Vaccination and Incidence computed per region. Both of the 2 dimensions can reflect hospital Saturation and define criteria according to which regions are labelled as either yellow, orange or red zones (Zona Gialla, Zona Arancione e Zona Rossa), in compliance with the most recent orders (D.L. 52/2021).

[R](#) [GitHub actions](#) [DataWrapper](#) **TECHNICAL SKILLS**

Coding	Python, R, SQL, HTML, CSS, JavaScript
Skills	Web scraping, API, Machine Learning
Data Science Frameworks	Sci-kit Learn, Keras, PySal, Tidyverse, Shiny, Hydra, Tidymodels
Database	PostgreSQL, PostGIS, MongoDB
Cloud	Linux Server, Docker, Amazon Web Services (AWS), Google Cloud Platform (GCP), DigitalOcean
Digital platforms	Google Analytics, Google AdWords, Facebook Ads
Web development	WordPress, Ecommerce
Others	Git, GitHub

 **COMMUNITY AND VOLUNTEER EXPERIENCE****BOARD MEMBER | DATA NETWORK**

2019

Data Network is a no-profit association that wants to bring in a unique environment students, professionals and companies that believe in the power of data and the change that could bring to modern society.

[website](#)**CONTRIBUTOR | OFFICINE ITALIA**

2020

Built together with other international colleagues a set of text mining tools (3 minimal web apps) analysing recovery fund plans. The objective was to measure the inclusion (via NLP) "youth" themes across EU countries.

[website](#) **LANGUAGES**

Italian	●	●	●	●	●
English	●	●	●	●	○
French	●	●	●	○	○

17 Mar 2022