

# Curriculum Vitae

## Oltion Preka, Ph.D.

### ESPERIENZA LAVORATIVA

#### Data Scientist

Nucleo di Valutazione e Verifica degli Investimenti Pubblici E-R, Regione Emilia-Romagna  
12/2020 – 10/2023

Come parte del team di Data Science, il mio compito principale era quello di aiutare le istituzioni locali nell'adozione di decisioni basate sui dati (*data-driven decision making*), contribuendo alla progettazione e l'implementazione di sistemi di monitoraggio di due programmi di finanziamento molto rilevanti per il territorio dell'Emilia-Romagna:

- (1) i Fondi di Programmazione dell'UE
- (2) il PNRR - Next Generation EU

Le mie responsabilità variavano ampiamente, spaziando dalla creazione di pipeline di dati alla creazione di dashboard in Power BI, a scrivere c.d. note per i *decision-makers*, e possono essere sintetizzate come segue:

- Identificazione dei dati provenienti da fonti interne (altre divisioni) e da fonti esterne (database nazionali o europei).
- Estrazione, pulizia e trasformazione dati (processi ETL) utilizzando il linguaggio di programmazione Python e/o lo strumento di PowerBI a secondo dalla convenienza in termini di efficienza.
- Contribuire alla progettazione di dashboard Power BI per monitorare i due programmi sopra menzionati o altri. Poiché il nucleo si occupa di compiti trasversali, tale processo ha coinvolto diversi altri attori di natura amministrativa al fine di meglio comprendere le loro esigenze informative e gli obiettivi di comunicazione. D'altra parte, il nostro team ha fornito una varietà di potenziali KPI e layout di Power BI.
- Costruzione e mantenimento di vari dashboard Power BI attraverso un approccio costruttivo cumulativo, una volta concordate le informazioni utili e come trasmetterle.
- Aggiornamento e modifica dei sistemi di monitoraggio man mano che le priorità informative cambiavano nel tempo.
- Partecipazione attiva a consultazioni con altri team, esplorando nuove opportunità di integrazione con altri sistemi e nuove fonti di dati.

#### Analista Statistico & Ricercatore

Dipartimento delle Scienze Politiche e Sociali, Alma Mater Studiorum - Università di Bologna  
10/2018 – 11/2020

Come unico analista quantitativo di un progetto di ricerca triennale (**Joint Efforts in the Defence Industry Project - JEDI**), portato avanti da tre università europee (Bologna, Groningen e Södertörn, Stoccolma), ero responsabile della mappatura dell'industria europea della difesa (dal processo di raccolta dati alla raccomandazione di politiche), che rappresentava uno dei principali risultati nel secondo del progetto. In particolare, le attività principali includevano:

- Esplorazione dei criteri per distinguere le aziende di difesa da quelle civili.
- Raccolta dati di tutte le aziende di difesa europee e delle loro principali caratteristiche, utilizzando anche le tecniche di **web scraping** in Python.

- Creazione di un database per archiviare i dati generati durante l'attività precedente.
- Manutenzione e aggiornamento dei database.
- Pulizia dati, trasformazioni e classificazione.
- Raccolta, manipolazione e analisi dei dati all'interno dell'industria europea della difesa.
- Visualizzazione dei cluster di difesa in Europa anche utilizzando mappe utilizzando il pacchetto **Folium**.
- Stesura di un rapporto sulla mappatura dell'industria della difesa in Europa.
- Infine, ho scritto il paper "*Sources of Strength: Mapping the European Defence Industry*", pubblicata all'interno di uno *special issue* della **rivista scientifica Defence Studies**, p. 531 - 560, Volume 23 (4), 2023, DOI: <https://doi.org/10.1080/14702436.2023.2277436>

## Data Scientist

Dipartimento delle Scienze Politiche e Sociali, Alma Mater Studiorum -Università di Bologna

01/06/2019 – 12/2019

Questo progetto aveva l'obiettivo di investigare come, a seguito delle Primavere Arabe, sono cambiate le relazioni tra i paesi del Medio Oriente e del Nord Africa (MENA) non solo all'interno della regione ma anche con le principali potenze globali (USA, Cina, Russia, Regno Unito, Francia, Germania e Italia). La ricerca è stata condotta adottando una metodologia quantitativa innovativa, che combinava l'analisi dei flussi commerciali tra i paesi con il livello dello sviluppo macro economico a livello del singolo paese, e i rapporti politici tra paesi (misurata attraverso i voti nell'ambito dell'ONU su questioni di rilevanza geostrategica internazionale, particolarmente legati al Medio Oriente). Tale approccio mirava a catturare il grado di distanza tra i paesi da una prospettiva sia economica che socio-politica.

Le principali attività includevano:

- Identificare, trovare ed estrarre dati da varie fonti affidabili.
- Pulire, trasformare e combinare dati garantendo la **coerenza tra dati** provenienti da diverse fonti.
- Condurre **cluster analysis** con l'obiettivo principale di gettare luce sulle relazioni tra i paesi.

## Docente universitario a contratto

Dipartimento delle Scienze Politiche e Sociali, Alma Mater Studiorum - Università di Bologna

10/2021 – Attuale

Ho insegnato il corso di laurea magistrale "**93031 - Data Journalism**" (4 CFU) per due anni accademici consecutivi, il **2021-2022** e il **2022-2023**, così come farò nell'attuale anno accademico **2023-2024**.

L'obiettivo del corso è di fornire gli studenti con esperienza pratica nell'applicare le tecniche principali di *data journalism*. Tuttavia, con l'avvento di **ChatGPT** l'anno scorso, il focus del corso si è spostato sulla consapevolezza degli studenti relativamente alle opportunità offerte e le sfide poste da ChatGPT, enfatizzando il suo potenziale ruolo che potrà svolgere nel giornalismo.

## Docente universitario a contratto

Dipartimento delle Scienze Politiche e Sociali, Alma Mater Studiorum - Università di Bologna

01/2019 – 06/2019

Ho tenuto il corso di laurea magistrale "**85008 - Big Data for the Social Sciences**" (8CFU) durante gli **anni accademici 2018-2019, 2019-2020, 2020-2021**.

Ho fornito un contributo significativo alla completa ri-disegnazione del corso, in particolare sostituendo il principale strumento tecnico, il software **KNIME**, con Python per i suoi vantaggi incomparabili rispetto a tutti gli altri linguaggi di programmazione. Il corso segue un approccio *bottom-up* basato su esempi del mondo reale che consente agli studenti senza nessuna esperienza pregressa di programmazione di imparare come sfruttare i Big Data per riconoscere patterns e acquisire conoscenze nel loro ambito di expertise.

Gli argomenti principali del corso includevano:

- Insegnamento dei concetti fondamentali del linguaggio di programmazione Python.
- Introduzione alle principali librerie per effettuare analisi quantitative, **Pandas** in particolare.
- Tecniche di *data visualization* utilizzando **matplotlib** e **seaborn**.
- Tecniche basi di analisi di testo (**Natural Language Processing - NLP**), ad esempio, tf-idf, bag-of-words, utilizzando il pacchetto **NLTK**.

Inoltre, ho assistito gli studenti nella progettazione della struttura del compito finale, aiutandoli a superare sia le sfide metodologiche che quelle relative alla programmazione.

### Analista Statistico (Statistical Analyst)

Food and Agriculture Organization of the United Nations (FAO)

06/2018 – 10/2018, Roma

Durante questo l'incarico, ho messo a disposizione di un progetto di ricerca - il cui obiettivo era quello di esplorare l'impatto delle migrazioni internazionali dalle aree rurali dei paesi dell'Asia centrale, in particolare Kirghizistan, Tagikistan e Armenia, verso la Federazione Russa - le mie competenze statistiche, sia teoriche che pratiche nell'utilizzo di strumenti di analisi statistica come Stata e Python. In stretta collaborazione con un'altra collega esperta in tematiche migratorie, abbiamo lavorato in modo coeso per massimizzare le sinergie tra le nostre competenze.

In particolare, le mie responsabilità includevano i seguenti compiti:

- Condurre un'ampia ricerca dei dati provenienti da cinque diverse indagini sui nuclei familiari raccolte da organizzazioni internazionali, ciascuna contenente oltre cento (100) variabili, per un totale di più di 600 variabili; identificazione delle variabili rilevanti cruciali per gli obiettivi del progetto.
- Garantire l'armonizzazione dei dati tra le variabili ottenute da diversi questionari, facilitando la loro comparabilità.
- Effettuare tutti i step necessari di manipolazione e trasformazione dati per migliorare la qualità degli stessi.
- Calcolare di indici statistici sintetici, in particolare il Punteggio di Diversità Alimentare dei Nuclei Familiari (HDDS).
- Condurre analisi statistiche descrittive.
- Esplorare l'impatto delle migrazioni rurali nei paesi di cui sopra secondo tre diverse dimensioni:
  - (1) sviluppo socio-demografico,
  - (2) sviluppo agricolo e sicurezza alimentare e
  - (3) protezione sociale.
- Inoltre, ho apportato significativi contributi alla componente di ricerca del progetto, allineandomi strettamente alle esigenze dello stesso.

### Consulente Data Scientist

Diversi clienti del settore privato

12/2017 – 06/2018

Ho aiutato piccole e medie imprese (PMI) a prendere decisioni basate sull'analisi dei dati, sfruttando il potenziale degli strumenti di programmazione, in particolare Python, e una moltitudine di pacchetti disponibili. Ho esplorato gli ultimi avanzamenti nel campo dell'Informatica, concentrandomi in particolare sulle tecniche di **Machine Learning (ML)** e **Deep Learning (DL)**.

In particolare, le attività chiave includevano:

- Effettuare attività di data mining e analisi di Big Data su dataset sia pubblici che privati, con

l'obiettivo di fare emergere regolarità nascoste, tendenze, e relazioni tra le variabili provenienti dai big data disponibili in una vasta gamma di formati.

- Utilizzo di algoritmi di ML per la segmentazione dei clienti attraverso tecniche di clustering.
- Applicazione di modelli statistici multivariati e analisi delle serie storiche per stimare le tendenze future di una o più variabili.
- Estrazione e archiviazione di dati da varie fonti, incluse le piattaforme dei social media (come Twitter e Facebook) e altri dataset pubblici.
- Conduzione di attività di web scraping per raccogliere dati dal web.
- Lavorare con diversi formati di dati (come .csv, .txt, .mat, .h5, .pkl, .npy, ecc.), nonché database relazionali (SQL e simili).
- Creazione di *data visualization* utilizzando librerie Python come Matplotlib, Seaborn e Bokeh, con l'obiettivo di fornire rappresentazioni visive chiare e approfondite dei dati.

## Policy Analyst

Food and Agriculture Organization of the United Nations (FAO)

08/2017 – 12/2017, Roma

Come membro chiave di un team composto da tre persone, ho svolto un ruolo cruciale in un progetto con obiettivi molteplici. Le mie principali responsabilità riguardavano i seguenti compiti:

- Fornire assistenza alle autorità della ex Repubblica Jugoslava di Macedonia, in particolare l'Autorità Nazionale Designata (NDA), nella preparazione di una proposta di finanziamento per il "*Readiness and Preparatory Support Programme*" nell'ambito di *Green Climate Fund* (CGF). Il mio contributo ha portato a garantire un finanziamento di \$300.000 per la Macedonia.
- Fornire supporto a rappresentanti di alto livello del governo macedone, in particolare del Ministero dell'Agricoltura, al fine di capire meglio le loro esigenze, risorse disponibili e aspettative riguardo alla collaborazione con la FAO all'interno del progetto.
- Stabilire linee di comunicazione efficaci con altre stakeholder già coinvolti nell'ambito delle politiche agricole nel paese, come la Banca Mondiale. Queste interazioni hanno facilitato le trattative tra il nostro team e l'ufficio regionale della Banca Mondiale, aprendo la strada a collaborazioni che si sono concretizzati nel futuro.

Inoltre, ho condotto:

- Analisi statistiche approfondite della produzione agricola e dei flussi commerciali internazionali.
- Calcolo di alcuni indicatori sviluppati dalla FAO volti a valutare il grado di apertura effettiva del mercato agricolo verso l'esterno. Tali analisi miravano a contribuire al disegno di politiche agricole allineate ai rigorosi requisiti del processo di adesione all'Unione Europea.

## Ricercatore Post-Doc

School of Agriculture & Food Science, University College Dublin (UCD)

08/2015 – 10/2016, Dublino (Irlanda)

Questo progetto di ricerca aveva l'obiettivo di esplorare la relazione causale tra innovazione e competitività nel settore agricolo in Irlanda, a livello micro economico, con un focus sulla performance del sistema di innovazione irlandese rispetto ai suoi principali concorrenti nel contesto internazionale. Sotto il profilo degli strumenti, **le analisi statistiche** e la stima dei vari **modelli econometrici** sono state condotte utilizzando Stata.

In particolare, le principali attività includevano:

- Approfondita conoscenza del concetto di innovazione nell'agricoltura, con un particolare attenzione sulle caratteristiche di un contesto favorevole affinché l'innovazione si verifichi; esplorazione del meccanismo di diffusione e degli effetti degli incentivi finanziari a livello micro delle singole aziende agricole;
- Analisi statistiche dei questionari European Union Farm Structure Survey (FSS) utilizzando il software Stata;
- Costruzione di indici compositi in grado di catturare gli aspetti innovativi all'interno del nucleo agricolo, individuare indicatori idonei e misurare il livello di innovazione agricola a livello micro;

- Applicazione di modelli econometrici e tecniche non parametriche per definire la relazione tra innovazione e competitività utilizzando il **software Stata**.

## Economista/ Ricercatore

Studiabo SRL

05/2015 – 08/2015

Ero responsabile della realizzazione di un progetto end-to-end (**dalla ricerca allo sviluppo del modello**) con l'obiettivo finale di sviluppare un modello in grado di prevedere le esportazioni dei prodotti 'Made in Italy' in futuro, adottando un approccio innovativo che andava oltre la considerazione dei fattori legati allo sviluppo economico.

In particolare, ero responsabile delle seguenti attività:

- Ricerca i fattori di natura non economica che possono impattare lo sviluppo economico futuro di un paese, evidenziando le differenze tra i paesi occidentali e quelli del resto del mondo, rivelatasi un aspetto fondamentale per le relazioni commerciali.
- Sviluppo di un modello statistico per prevedere le future esportazioni verso la maggior parte dei paesi del mondo (compresi quelli asiatici), che tenesse conto di fattori sociodemografici, geografici, culturali e religiosi, oltre ai tradizionali fattori economici.
- Raccolta dati da diverse fonti di dati pubblicamente disponibili.
- Pulizia e trasformazione dei dati, poiché i dati sono presenti in una vasta gamma di formati, utilizzando strumenti di business intelligence come PowerBI.
- Esplorazione dei dati e visualizzazione dei dati per comprendere meglio le principali statistiche descrittive di ciascuna variabile.
- Manipolazione dei dati.
- Sviluppo del modello economico finale dopo aver testato una vasta gamma di modelli differenti.

## Docente universitario

Facoltà di Economia e Agro-business, Università Agricola di Tirana

08/2013 – 09/2014, Tirane (Albania)

Le principali attività svolte includevano:

- Insegnamento del Corso "**Introduzione alla Statistica**" (8-CFU).
- Svolgimento di attività di ricerca.
- Partecipazione a compiti amministrativi (ad es. Consiglio del Dipartimento).

## ISTRUZIONE E FORMAZIONE

### PhD in Economia e Statistica Agroalimentare

Facoltà di Scienze Statistiche, Alma Mater Studiorum - Università di Bologna

12/2009 – 06/2014

Area di ricerca: **Valutazione delle politiche di sostegno finanziario al settore agricolo**

Titolo tesi finale: "***Evaluating the impact of grants in the agricultural sector in Albania: a counterfactual approach using propensity score matching***"

### Laurea Specialistica in Economia e Diritto (summa cum laude)

Facoltà di Economia, Alma Mater Studiorum - Università di Bologna

08/2006 – 12/2008

### Laurea Triennale in Economia e Diritto (110/110)

Facoltà di Economia, Alma Mater Studiorum - Università di Bologna

07/2003 – 11/2006

## Altre Competenze

### Competenze Linguistiche

Lingua Madre	Albanese				
Italiano	Pari a madrelingua				
	Ascolto	Lettura	Interazione orale	Produzione orale	Scrittura
Inglese	C2	C2	C2	C2	C2
Spagnolo	B1	B2	A1	A2	A1

### Competenze Digitali

Durante una lunga esperienza formativa e lavorativa che si estende dal dottorato di ricerca più di dieci anni fa ad oggi ho sviluppato le seguenti capacità tecniche e di programmazione principali.

Ottima conoscenza del **linguaggio di programmazione Python**, in particolare, delle seguenti librerie:

- **Pandas** (analisi data)
- **Matplotlib, Bokeh, Seaborn, Folium** (visualizzazione dati)
- **NumPy, SciPy** (calcoli scientifici & analisi statistiche)
- **Scrapy** (web scraping)
- **SpaCy, NLTK** (text mining)
- **Scikit-learn, Sklearn\_pandas** (Machine Learning)
- **PyTorch** (Deep Learning Framework)
- **Transformers, Datasets, HF Hub** (Hugging Face)

Ottima conoscenza dei seguenti strumenti DevOps:

- **GIT**

Altri strumenti:

- Linux (Sistema operativo *open source*)
- Anaconda (ambiente di lavoro Python)
- Jupyter Notebooks
- VS Code

Inoltre:

Ottima conoscenza dello strumento di business intelligence **Power BI**;

Ottima conoscenza di programmazione del software statistico **STATA**;

Ottima conoscenza del linguaggio **SQL**;

Buona conoscenza del software statistico **SPSS**.

### Corsi Rilevanti

#### **Practical Deep Learning for Coders - Part 2**

collaborazione tra **l'Università di Queensland con Fast.ai**

06/2022 – 11/2022, Online

## Practical Deep Learning for Coders - Part 1

collaborazione tra l'Università di Queensland con Fast.ai

03/2022 – 06/2022, Online

## Third International School on Open Science Cloud - SOSOC 2019

organizzato dall'Istituto Nazionale di Fisica Nucleare (INFN), Università di Perugia e Università di Bologna

<https://agenda.infn.it/event/19049/>

15/09/2019 – 30/09/2019, Bologna

## Partecipazioni a corsi brevi ed eventi

- “Python Conference Ireland (PYCONIE)”, Novembre (2016) - Dublino (Irlanda)
- “OPENLOC School on Fundamentals of Economic Policy Evaluation”, organizzato da una collaborazione dall'Università di Trento e l'Università di Bologna, Ottobre (2011) - Bologna
- “Lobbying at the European Institutions” - Winter School organizzato da Eqo presso il Parlamento Europeo, Dicembre (2009) - Bruxelles (Belgio)

## Reti e associazioni di appartenenza

Membro del **Centro di Computational Social Science (CSSC)**, Dipartimento delle Scienze Politiche e Sociali, Università di Bologna

[Computational Social Science Center - \(CSSC\) \(unibo.it\)](https://www.cssc.unibo.it/)

06/2021 - Bologna

CSSC - Il Centro per Computational Social Science (CssC) si propone di promuovere l'impiego di metodi, strumenti e tecniche delle scienze computazionali per ampliare la capacità di analisi delle scienze sociali. Il Centro è ospitato dal Dipartimento di Scienze Politiche e Sociali (DSPS) ma ha fra i suoi membri colleghi provenienti da altri Dipartimenti dell'Ateneo.

## **Socio Fondatore di StatisticaMente** - Associazione di Promozione Sociale

(P. IVA 91427670376)

05/2020 – Attuale, Bologna

Sono uno dei 6 soci fondatori di StatisticaMente, un'associazione di promozione sociale, i cui obiettivi principali sono di promuovere la diffusione della **statistica applicata** e migliorare la conoscenza dei metodi statistici nell'ambito dei ricercatori medici. Il nostro associato tipico si trova tra i medici, specialmente tra coloro che sono coinvolti nella comprensione e nella scrittura di articoli scientifici. Oltre alla consulenza statistica, l'associazione offre corsi brevi rivolti alla categoria dei medici. Tali corsi sono stati frequentati da diverse decine di partecipanti provenienti da tutta l'Italia, nonché da una vasta gamma di specializzazioni mediche.

## Altri Progetti

### Artificial Intelligence (AI) wine similarity

05/2020 – 01/2021

- [www.winesimilarity.com](http://www.winesimilarity.com)