

## Curriculum vitae

### INFORMAZIONI PERSONALI

Daniele Sartiano

### ESPERIENZA PROFESSIONALE

Mar. 21 – Freelancer

Coordinamento tecnico di team per lo sviluppo di applicazioni basate su intelligenza artificiale. Disegno e sviluppo di applicativi software con utilizzo di Intelligenza Artificiale, con focus sulla parte di backend, microservizi e data analysis.

Attività o settore Ricerca e Sviluppo

Apr. 21 – Apr. 22

**Incarico di collaborazione professionale: "Studio ed elaborazione di strumenti e procedure per l'analisi e la valutazione dell'accuratezza dei dati associati ai nomi a dominio .it"**

Ministero dell'Università e della Ricerca – Consiglio Nazionale delle Ricerche – Istituto di Informatica e Telematica, Pisa (Italia)

Progettazione e sviluppo di un sistema di analisi della correttezza dei dati dei registranti dei nomi a dominio del "Registro .it". Il software, in parte già realizzato durante l'assegno di ricerca nel precedente periodo, è stato consolidato e sono state aggiunte feature per l'analisi dei dati e per la gestione delle anomalie rilevate, comprendenti sistemi basati su autoencoder. È stato effettuato un refactoring del modello dei dati, al fine di incrementare le performance del sistema, inoltre è stata progettata l'interfaccia per l'integrazione dell'applicazione verso altri sistemi del Registro .it.

Attività o settore Ricerca e Sviluppo

Mar. 19 – Mar. 21

**Assegno di Ricerca: "Metodologie e algoritmi di machine learning per l'analisi e la classificazione di contenuti web"**

Ministero dell'Università e della Ricerca – Consiglio Nazionale delle Ricerche – Istituto di Informatica e Telematica, Pisa (Italia)

Progettazione e realizzazione di strumenti per la classificazione automatica di pagine web e la data analysis dei dati del "Registro .it".

Gli strumenti sono sviluppati con tecnologie allo stato dell'arte, in particolare sono basati su *deep learning*, sia *unsupervised* che *supervised*, utilizzando librerie allo stato dell'arte: *Keras*, *Tensorflow* e *Pytorch*.

Progettazione e sviluppo di un sistema di analisi della correttezza dei dati dei registranti dei nomi a dominio del "Registro .it".

Il sistema si basa su un'architettura a microservizi, su *docker compose*, in cui sono stati disegnati e implementati moduli per la lettura e l'aggregazione dei dati dal database del registro, la classificazione di questi al fine di trovare inconsistenze, un'interfaccia web per la lettura delle analisi effettuate e le eventuali annotazioni fatte dal personale, un'interfaccia REST per l'integrazione con altri servizi. Il sistema intelligente è stato definito progettando dei classificatori basati su *Deep Learning*, in

20/04/2022

particolare con *Keras* e *Tensorflow*.

Progettazione e sviluppo di un classificatore supervisionato basato su Gradient Boosting in grado di identificare traffico *DoH* (*Dns Over Https*) su traffico criptato.

Analisi del traffico di rete con identificazione di nuove euristiche per identificare eventuali anomalie su traffico criptato IoT, con pubblicazione dei seguenti:

L. Deri and D. Sartiano  
Monitoring IoT Encrypted Traffic with Deep Packet Inspection and Statistical Analysis [ IEEE ]  
Proceedings of 15th International Conference for Internet Technology and Secured Transactions (ICITST-2020), 2020.

L. Deri and D. Sartiano  
Using DPI and Statistical Analysis in Encrypted Network Traffic Monitoring  
International Journal for Information Security Research, Volume 10 Issue 1, 2020.

Partecipazione alla competizione internazionale "Suspicious Network Event Recognition", nell'ambito della quale ho implementato due classificatori per il riconoscimento di eventi sospetti di rete: il primo basato su Gradient Boosting, in particolare su XGBoost; il secondo basato su Deep Learning. Il lavoro è stato descritto nell'articolo "Suspicious Network Event Recognition Leveraging on Machine Learning", che è stato pubblicato e presentato alla conferenza IEEE Big Data 2019.

**Attività o settore** Ricerca e Sviluppo

Sett. 19 – Sett 20

**Incarico di lavoro autonomo: "Progettazione e sviluppo di un sistema di detection di anomalie sul traffico di rete, di fonti di malware e di siti malevoli tramite tecniche di Deep Learning"**

Ministero dell'Università e della Ricerca – Università di Pisa – Dipartimento di Informatica

Progettazione e sviluppo di un sistema per il rilevamento di anomalie sul traffico di rete attraverso l'utilizzo di Deep Learning.

Partecipazione alla competizione internazionale "IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies", in cui abbiamo sviluppato un parser a dipendenze basato su Deep Learning. Il lavoro è stato descritto nell'articolo: "Linear Neural Parsing and Hybrid Enhancement for Enhanced Universal Dependencies", che è stato pubblicato e presentato alla conferenza "IWPT 2020: the 16th International Conference on Parsing Technologies".

**Attività o settore** Ricerca e Sviluppo

Mar. 19 – Giu. 20

**Software Architect / Data Scientist**

Dataninja S.r.l.s. / Gruppo 24 ore

Collaborazione per lo sviluppo di un applicativo in grado di generare podcast a partire da dati strutturati.

Il progetto è stato finanziato da Google DNI (Digital News Initiative) ed è stato sottomesso dal Gruppo 24 ore, il Dipartimento di Informatica dell'Università di Pisa e Dataninja srls, una startup che si occupa di analisi di dati.

Il mio ruolo all'interno del progetto è quello di architetto del software, sviluppatore software e data scientist.

L'applicazione è stata disegnata con un'architettura a micro-servizi con container *Docker*. I micro-servizi comunicano attraverso code (*RabbitMQ*), il *frontend* comunica con il *backend* attraverso *websocket* e l'analisi dei dati viene effettuata tramite *Spark*.

Come sviluppatore software mi occupo dello sviluppo del *backend*.

Come data scientist mi occupo della costruzione di prototipi per l'analisi del testo, come ad esempio lo sviluppo di un prototipo per la correzione automatica del testo prodotto per la generazione dei podcast.

Ref: <https://newsinitiative.withgoogle.com/dnifund/dni-projects/ai-anchor-round-5/>

20/04/2022.

## Curriculum vitae

Mar. 18–Mar. 19

### Full stack developer

Camlin Technologies, Pisa (Italia)  
[www.camlingroup.com](http://www.camlingroup.com)

Disegno e implementazione di applicazioni distribuite nel campo dell'e-health. Nello specifico ho lavorato ad un sistema per la riabilitazione domiciliare di pazienti post ictus (<https://www.camlingroup.com/arc-intellicare>).

Dopo essermi occupato dello sviluppo di alcune feature sulle applicazioni Android (scritte utilizzando *Android Architecture Components*) in Java, mi sono occupato della parte di backend, scritta in *Django* e *Django Rest Framework*, su architettura a container (Docker). Il sistema mette a disposizione dei pazienti in terapia post ictus, la possibilità di effettuare da casa, anche in assenza di rete, gli esercizi fisioterapici prescritti dal fisioterapista, dandogli un feedback sull'esecuzione sfruttando librerie di machine learning sviluppate ad hoc.

Durante lo sviluppo del software abbiamo utilizzato la metodologia *scrum*.

Attività o settore e-health

Mar. 15–Feb. 18

### Borsa di Ricerca: Metodi e strumenti per la classificazione di pagine web

Ministero dell'Università e della Ricerca – Università di Pisa – Dipartimento di Informatica

Disegno e implementazione di strumenti per l'analisi del linguaggio naturale per la classificazione automatica di pagine web.

Il mio ruolo in questo progetto è stato quello portare il contributo del gruppo di Human Language Technologies del Dipartimento di Informatica dell'Università di Pisa, costruendo due classificatori per la classificazione automatica di pagine web, il primo attraverso *Support Vector Machine*, il secondo attraverso tecniche deep learning utilizzando *word embeddings* e *convolutional neural network* (<https://github.com/daniele-sartano/DeepWebClassifier>).

Demo: Food in the net Observatory <http://www.foodinthenet.it/>

E' stato inoltre implementato un classificatore web basato su classificazione per contesto, in cui il core è stato scritto in Java e basato su map reduce, con elaborazione distribuita basata su Hadoop e database distribuito basato su HDFS.

Nello stesso periodo con il gruppo di Human Language Technologies del Dipartimento di Informatica dell'Università di Pisa abbiamo partecipato alla competizione internazionali Evalita 2016 (<http://www.evalita.it/2016>) in cui siamo arrivati primi ai task di "Named Entity Recognition and Linking in Italian Tweets" e "SENTiment POLarity Classification".

Mi sono inoltre occupato dell'amministrazione delle macchine del gruppo, basate su sistemi Linux.

Feb. 12–Feb. 18

### Co-fondatore

BioCare Provider srl, Pisa (Italia)  
[biocareprovider.com/it/home/](http://biocareprovider.com/it/home/)

Nel Febbraio 2013, io e alcuni miei colleghi, abbiamo fondato BioCare Provider, una startup innovativa spinoff dell'Università di Pisa e della Scuola Superiore Sant'Anna. La mission di BioCare era di mettere le nuove tecnologie informatiche a disposizione di malati ed anziani, al fine di favorire la comunicazione con il proprio medico curante ed i propri familiari, e supportare al tempo stesso sia l'assistenza domiciliare che la ricerca clinica.

Nel febbraio 2018 la società è stata acquisita da Camlin Technologies, che ha acquisito il 100% delle quote di BioCare.

Oltre a socio fondatore, i miei ruoli in BioCare erano quelli di Project Manager, System Architect e sviluppatore.

Come Project Manager ho lavorato al design e la gestione di progetti IT, coordinando i progetti con aziende esterne e gestendo le relazioni con i clienti.

Come System Architect e sviluppatore ho progettato e sviluppato web dashboard in ambito sanitario,

20/04/2022



affrontando la progettazione e lo sviluppo del modello dei dati, del backend, del frontend e del deployment, come ad esempio il sistema gestione per Clinica Mobile nel Mondo (<https://www.clinicamobile.com/>), oppure i portali per l'Agenzia Regionale in Sanità (<https://proter.ars.toscana.it/>, <https://esiti.ars.toscana.it/>, <https://prolea.ars.toscana.it/>). Ho progettato e sviluppato parte del prodotto principale di BioCare (<https://algos.drdrin.it/>), un sistema per mettere in comunicazioni pazienti e medici, facilitando la gestione delle terapie farmaceutiche. In particolare, oltre al design delle API rest, ho lavorato sia al backend del sistema, sviluppato con Django/Django Rest Framework, sia allo sviluppo del portale web (HTML/Javascript), sia alle applicazioni ibride attraverso il framework Ionic.

Attività o settore e-health

#### Set. 16–Gen. 17 **Analisi automatica di traffico di rete criptato**

Disegno e implementazione di uno strumento per il riconoscimento di specifiche sequenze di pacchetti utilizzando tecniche di Deep Learning.

Attività o settore Network Monitoring

#### Ago. 13–Feb. 15 **Assegno di Ricerca: "Estrazione ed analisi di informazioni sanitarie espresse in linguaggio naturale"**

Ministero dell'Università e della Ricerca – Università di Pisa – Dipartimento di Informatica, Pisa (Italia)

Progettazione e realizzazione di strumenti in grado di estrarre informazioni di tipo medico da referti clinici espressi in linguaggio naturale.

Gli strumenti fanno parte di una completa pipeline linguistica, che utilizzano tecniche di machine learning per l'apprendimento di tipo *supervised* e *unsupervised*.

Stesura dei deliverable e rendicontazioni relative al progetto di ricerca.

Partecipazione a competizioni internazionali per il riconoscimento di entità nel dominio clinico.

Amministratore delle macchine del gruppo di Human Language Technologies, basate su sistemi Linux.

Attività o settore Ricerca e Innovazione in Sanità

#### Feb. 12–Lug. 13 **Assegno di Ricerca: "Tecniche di Traduzione Automatica Statistica basate su sintassi"**

Ministero dell'Università e della Ricerca – Università di Pisa – Dipartimento di Informatica, Pisa (Italia)

Creazioni di corpus paralleli, alcuni ottenuti direttamente dal web estraendo contenuti dalle pagine html.

Configurazione e utilizzo del software Moses (<http://www.statmt.org/moses/>), dei vari tools per il training/tuning dei Language Model e dei modelli utilizzati dai Moses per la traduzione.

Messa in opera di servizi web per la traduzione automatica, utilizzando le API messe a disposizione da Moses.

Attività o settore Ricerca su traduzione automatica

#### Apr. 12–Lug. 12 **Collaborazione a Progetto**

Tiscali Italia S.p.A., Pisa (Italia)

Il progetto di cui mi sono occupato è la realizzazione di un modulo di *Web Query Classification*, implementato attraverso l'utilizzo di tecniche di Machine Learning:

- Raccolta dei dati attraverso lo sviluppo di un'interfaccia web messa a disposizione dei dipendenti dell'azienda, scritta in Python (<http://www.tomadoweb.org/>).
- Definizione delle Features e successiva Features Selection. Le *feature* sono state selezionate

20/04/2022

## Curriculum vitae

anche mediante l'utilizzo di strumenti e tecniche di elaborazione del linguaggio naturale.

- Progettazione del classificatore, implementato inizialmente in *Python*, utilizzando *libsvm* (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).
- Successiva implementazione del classificatore in C++ ed integrazione di esso in un web server efficiente (<http://fallabs.com/kyototycoon/>) per poterlo interrogare tramite servizi REST.

**Attività o settore** Società di Telecomunicazioni

Apr. 07–Set. 07

### Sviluppatore Software

Eudata s.r.l., Milano (Italia)

In questa Società ho iniziato effettuando uno stage, dove ho appreso le nozioni fondamentali nell'area CTI, tra cui i sistemi Genesys, Cisco e il protocollo VOIP.

Dopodiché ho incominciato a seguire due progetti da sviluppatore di software:

- Un software per la gestione di una periferica che funge da Dosimetro Acustico, da collegare tra la cuffia e l'apparecchio telefonico. Il progetto comprende una parte client, sviluppata in C++ con l'integrazione di una DLL implementata ad hoc, e una parte server, sviluppata in JSP + MySQL.
- Una Barra Telefonica, la quale si interfaccia ad un server *Asterisk*. Questo programma è stato sviluppato in *JAVA/JSP* e *Javascript*, con tecnologia Ajax e si appoggia ad un DBMS MySQL. La barra telefonica comunica con il server *Asterisk*, fungendo quindi da client per esso.

**Attività o settore** Società di consulenza informatica che lavora nell'area CTI (Computer Telephony Integration)

## ISTRUZIONE E FORMAZIONE

Set. 07–Apr. 11

### Dottore Magistrale in Tecnologie Informatiche con votazione 110/110.

Università di Pisa – Facoltà di Scienze MM. FF. NN., Pisa (Italia)

Titolo tesi: "Disegno e implementazione di un framework per la misura qualitativa dei servizi di rete".

La tesi propone un'architettura per l'analisi qualitativa dei servizi di rete, mediante la misura permanente della loro qualità, così come percepita da parte degli utenti. Nello specifico è stata misurata la qualità dei servizi su HTTP, in particolare della navigazione web e dello streaming on demand.

Relatore: Luca Deri

Controrelatore: Antonio Brogi

Ott. 03–Feb. 07

### Dottore in Informatica con votazione 110/110 e Lode

Università dell'Insubria – Facoltà di Scienze MM. FF. NN., Varese (Italia)

Titolo tesi: "Valutazione del rischio di attacco ad una rete dall'esterno: il caso del Dipartimento di Informatica e Comunicazione".

In questa tesi ho analizzato la rete del dipartimento di Informatica e Comunicazione di Varese, effettuando una stima del rischio di questa.

Relatore: Marco Benini

Set. 95–Lug. 01

### Diploma di Perito in Elettrotecnica e Automazione con votazione 77/100

Istituto Tecnico Industriale Statale di Gallarate, Gallarate (Italia)

## COMPETENZE PERSONALI

Lingua madre	italiano				
Lingue straniere	COMPRESIONE		PARLATO	PRODUZIONE SCRITTA	
	Ascolto	Lettura	Interazione	Produzione orale	
inglese	C1	C1	C1	C1	C1

Livelli: A1 e A2: Utente base - B1 e B2: Utente autonomo - C1 e C2: Utente avanzato  
Quadro Comune Europeo di Riferimento delle Lingue

Corso di Inglese effettuato da Ottobre 2019 a Maggio 2020 con attestazione di livello ottenuto "C1 – Effective Operational Efficiency" ottenuta il 25/05/2020 del CEFR (*Common European Framework of Reference*).

#### Competenze digitali

##### **SISTEMI OPERATIVI:**

Linux, Windows.

##### **DBMS:**

PostgreSQL, MySQL, SQLite, Hbase.

##### **LINGUAGGI:**

Python (Cython), Javascript/Typescript, Java, C/C++, HTML, XML, CSS, SQL, Bash Scripting.

##### **METODOLOGIE DI PROGRAMMAZIONE:**

Strutturata, Object Oriented, Map Reduce.

##### **ELABORAZIONE DEL LINGUAGGIO NATURALE:**

Strumenti per l'elaborazione del linguaggio naturale, quali Sentence Splitter, Tokenizer, Part-Of-Speech Tagger, Morphological Analyzers and Lemmatizers, Named-Entity Tagger, Supersense Tagger (Wordnet Supersense), Dependency Parser, Word Embeddings, Transformers.

##### **MACHINE LEARNING:**

Deep Learning (Keras, Tensorflow, Pytorch), XGBoost, Scikit Learn, Support Vector Machine (libsvm, liblinear).

##### **DATA ANALYSIS:**

Spark, Hadoop, Pandas.

##### **CONTAINER:**

Docker, Docker Compose.

##### **SVILUPPO MOBILE:**

Android, Ionic 2.

##### **SVILUPPO WEB:**

Django, Flask, FastAPI, Tomado, HTML, CSS, JQuery, Angular JS, Java Server Page, Spring MVC.

##### **DEVELOPMENT TOOLS:**

Emacs, PyCharm, VS Code, WebStorm, Eclipse, Aptana Studio, Netbeans, Spring Tool Suite, Visual

## Curriculum vitae

Studio.

### Progetti personali

#### penelope-bot

<https://github.com/daniele-sartiano/penelope-bot>

Disegno e implementazione di un crawler distribuito e asincrono scritto in C++. Il sistema è basato su docker compose, utilizza nats.io per lo scambio di messaggi tra i container e il database distribuito Scylla db per il salvataggio dei dati.

Linux, Windows.

#### DiaParser

<https://github.com/Unipisa/diaparser>

Sviluppo e manutenzione di DiaParser, un parser a dipendenze, il quale costruisce degli alberi a dipendenze a partire dal testo dato in input, in modo tale da poter sfruttare le informazioni semantiche per task di *Natural Language Processing*.

Il parser è scritto in *PyTorch* e sfrutta i transformers per ottenere risultati allo stato dell'arte.

### Patente di guida

A, B

## ULTERIORI INFORMAZIONI

### Attività formativa

- Dal 2013 aiuto il docente del corso di *Natural Language Processing*, il Professor Giuseppe Attardi, del Dipartimento di Informatica dell'Università di Pisa, nel seguire gli studenti per lo sviluppo dei progetti d'esame e per le loro tesi di laurea triennale e magistrale.
- 25 Marzo 2019 – "Il pensiero computazionale – Tides Google" – Ho tenuto un seminario insegnando i principi del linguaggio di programmazione *Python* e illustrando un esempio di utilizzo del *Reinforcement Learning*.

### Pubblicazioni

- L. Deri and D. Sartiano. "Monitoring IoT Encrypted Traffic with Deep Packet Inspection and Statistical Analysis". *Proceedings of 15th International Conference for Internet Technology and Secured Transactions (ICITST-2020)*, 2020.
- L. Deri and D. Sartiano. "Using DPI and Statistical Analysis in Encrypted Network Traffic Monitoring". *International Journal for Information Security Research, Volume 10 Issue 1*, 2020.
- Attardi, Giuseppe, Daniele Sartiano, and Maria Simi. "Linear Neural Parsing and Hybrid Enhancement for Enhanced Universal Dependencies." *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*. 2020.
- D. Sartiano, G. Attardi, L. Deri and M. Martinelli, "Suspicious Network Event Recognition Leveraging on Machine Learning," *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 2019, pp. 5915-5920, doi: 10.1109/BigData47090.2019.9006178.
- Sartiano, Daniele, and Giuseppe Attardi. "UniPI at DMD-2018 Shared Task: Bidirectional LSTM models for DGA classification." *Sixth International Symposium on Security in Computing and Communications*. Springer, 2018.
- Attardi, G., Sartiano, D., Simi, M., Sucameli, I., & Pontecorvo, L. B. (2016). Using Embeddings for Both Entity Recognition and Linking in Tweets.
- Attardi, G., Sartiano, D., Alzetta, C., Semplici, F., & Pontecorvo, L. B. Convolutional Neural Networks for Sentiment Analysis on Italian Tweets.
- Attardi, Giuseppe, and Daniele Sartiano. "UniPI at SemEval-2016 Task 4: Convolutional Neural Networks for Sentiment Classification." *Proceedings of SemEval* (2016): 220-224.
- Deri, Luca, et al. "Implementing Web Classification for TLDs." *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. Vol. 1. IEEE, 2015.
- Attardi, Giuseppe, Vittoria Cozza, and Daniele Sartiano. "Detecting the scope of negations in

2-11-2022