

## INFORMAZIONI PERSONALI

Daniele Sartiano

## ESPERIENZA PROFESSIONALE

Mag. 22 – In Corso

### Incarico di collaborazione professionale: “Studio e 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)

Incarico professionale con protocollo: 0003138/2022 del 05/05/2022 nell'ambito del progetto DIT.AD005.010 - P0000127 “Cybersecurity”.

Progettazione e sviluppo di nuove funzionalità ad un sistema di analisi della correttezza dei dati dei registranti dei nomi a dominio del Registro .it. Il sistema attraverso euristiche, algoritmi di deep learning e servizi esterni, identifica eventuali inesattezze e incoerenze tra i dati, tra cui i registranti aventi domini utilizzati per scopi malevoli. Il sistema è stato scritto in Python, il modello dei dati è relazionale su PostgreSQL, le reti neurali sono scritte in Keras e Tensorflow e l'interfaccia web è scritta in HTML/Json, la quale comunica via REST con il backend, sono inoltre esposte delle API per l'integrazione con altri servizi.

Al software, in parte già realizzato durante le collaborazioni precedenti, è stato integrato la comunicazione sicura con altre piattaforme del Registro .it, inoltre è stato predisposto il deploy per il testing su piattaforma distribuita basata su Docker Swarm.

È in corso la progettazione per l'analisi dei nomi a dominio, come ad esempio il riconoscimento di domini generati attraverso algoritmi DGA, e cyber squatting.

**Attività o settore** Progettazione e sviluppo software, Natural Language Processing, Machine Learning e Deep Learning, Cybersecurity, Data Science

Gen. 22 – Giu. 22

### Incarico di collaborazione professionale: Software Architect / Data Scientist

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

Collaborazione per lo sviluppo di nuove funzionalità di un sistema di generazione di testi a partire da dati strutturati. In particolare, è stato costruito un verticale per la generazione di pagine web, con contenuti in linguaggio naturale. Queste pagine web contengono informazioni relative a tutte le aziende italiane quotate in borsa, con ad esempio informazioni sui titoli azionari, compagine aziendale, competitor, ecc. L'applicazione ha 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 *PySpark*. Io mi sono occupato principalmente della parte di backend, andando a modificare il modello dei dati, il sistema per la ricezione dei dati strutturati da multiple fonti, il generatore di testi.

**Attività o settore** Progettazione e sviluppo software, Natural Language Processing, Calcolo distribuito

Set. 21 – Giu. 22

### Incarico di collaborazione professionale: “Ricerca e sviluppo di applicazioni basate su Machine Learning”

Fluel S.r.l.s.

Incarico di collaborazione professionale per lo sviluppo di applicazioni basate su machine learning, tra cui sistemi per lead generation, classificazione di immobili e identificazione di obsolescenza di magazzino. Per tutti i progetti vi è stata una fase di analisi di analisi dei requisiti, di disegno dell'architettura del software, di analisi dei dati a disposizione, *preprocessing* di quest'ultimi, sperimentazione per l'identificazione dei migliori modelli di *Machine Learning* da utilizzare ed implementazione della pipeline completa per l'analisi dei dati, incluso l'implementazione di *web*

services per la fruizione dei servizi implementati.

**Attività o settore** Progettazione e sviluppo software, Machine Learning, Clustering, Data Science

**Set. 21 – Dic. 21** **Incarico di collaborazione professionale: “Direzione tecnica”**

S.AI – Synthema Artificial Intelligence Srl

Direzione tecnica del team di sviluppo, responsabile della pianificazione delle risorse e degli obiettivi aziendali.

Analisi dei requisiti per la progettazione e lo sviluppo di applicazioni basate su Natural Language Processing e Deep Learning.

Analisi del software aziendale e definizione dei requisiti per il rinnovo di esso, attraverso tecnologie allo stato dell'arte.

Progettazione di un sistema che effettua analisi aziendali attraverso l'utilizzo di tecniche di Natural Language Processing e l'integrazione verso servizi esterni.

**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)

Incarico professionale con protocollo: 0002652/2021 del 15/04/2021 nell'ambito del progetto: “Sviluppo Applicazioni Registro”.

Progettazione e sviluppo di un sistema di analisi della correttezza dei dati dei registranti dei nomi a dominio del “Registro .it”. Questo sistema dà la possibilità di individuare inesattezze e incoerenze tra i dati, tra cui eventuali registranti con domini utilizzati per scopi malevoli.

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 .it, l'analisi di questi al fine di trovare anomalie, 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 effettua l'analisi dei dati utilizzando euristiche, algoritmi di machine learning, tecniche di natural language processing e servizi esterni, in particolare il sistema intelligente è stato definito progettando delle reti neurali basate su Deep Learning, scritte tramite Keras e Tensorflow. Queste reti prendono in input dei dati non strutturati e sono allenate per identificare delle anomalie in essi.

Il sistema è stato scritto in Python, il modello dei dati è relazionale su PostgreSQL, l'interfaccia web è scritta in HTML/Json, la quale comunica via REST con il backend, sono inoltre esposte delle API per l'integrazione con altri servizi.

La piattaforma è stata presentata durante il CyberSecurity Day dell'Internet Festival di Pisa e al CIR - Comitato consultivo di Indirizzo del Registro .it.

**Attività o settore** Ricerca e Sviluppo

**Gen. 21 – Lug. 21** **Incarico di collaborazione professionale: consulenza su NLP**

Dataninja S.r.l.s.

Consulenza per la costruzione di una pipeline di analisi linguistica che utilizza Deep Learning e la tecnologia allo stato dell'arte.

La consulenza si è poi focalizzata sulla generazione di testi a partire da dati strutturati attraverso tecniche di machine learning e in particolare Deep Learning.

**Attività o settore** Consulenza tecnica su Natural Language Processing, Machine Learning e Deep Learning

## Curriculum vitae

- 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)
- Conferimento assegno di ricerca con protocollo 0002450/2019 nell’ambito del programma di ricerca “MoRSe – Monitoring Registrar Services platform” per la tematica “Metodologie e algoritmi di machine learning per l’analisi e la classificazione di contenuti web”.
- Protocollo rinnovo assegno di ricerca: protocollo 0002209/2020.
- Progettazione e sviluppo di un sistema per la classificazione di pagine web attraverso tecniche di Natural Language Processing basate su Deep Learning, in cui le feature calcolate sono state estratte dal contenuto delle pagine web e dal nome a dominio relativo.
- Progettazione e sviluppo di un prototipo per domain suggestion, il quale sfrutta tecniche di natural language processing per suggerire dei nomi a domini dato uno in input. In quest’ambito sono stati effettuati degli esperimenti per poter fornire dei suggerimenti intelligenti, sfruttando tecniche di machine learning, tra cui l’utilizzo di word embeddings per estrarre termini simili.
- Testing e debugging della piattaforma MoRSe (Monitoring Registrar Services), progetto finanziato nell’ambito del comitato di indirizzo del Registro .it. MoRSe è una piattaforma di monitoraggio in cloud, che offre ai Registrar la possibilità di verificare la disponibilità e la sicurezza dei propri servizi. I Registrar, attraverso MoRSe, hanno la possibilità di individuare problemi di rete, come ad esempio una bassa latenza, modifiche al routing BGP e di identificare le eventuali vulnerabilità delle proprie macchine.
- Progettazione e sviluppo di un classificatore binario supervisionato basato su Gradient Boosting in grado di identificare traffico DoH (Dns Over Https) su traffico criptato. Il classificatore prende in input flussi di rete e relativi metadati estratti da un software di Deep Packet Inspection, in particolare nDPI.
- Il codice sorgente è open source e disponibile online: <https://github.com/daniele-sartiano/doh>.
- I risultati sono stati presentati nel meeting “Encrypted DNS Policy and Technical Considerations Call” del 9 Novembre 2020.
- Analisi del traffico di rete con identificazione di nuove euristiche per rilevare 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 al Meeting di Centr a Leuven in cui si è discusso con gli altri partecipanti di come affrontare problemi come l’individuazione di fake webshop, metodi per il clustering di siti web, domain name suggestion e problemi rilevati sulla registrazione dei domini.
- Partecipazione al meeting di R&D a Pisa, in cui i partecipanti hanno mostrato le attuali attività di ricerca e sviluppo, in cui è stata presentata una demo di MoRSe. Nel secondo giorno ho partecipato ad un tutorial su tecniche di Natural Language Processing per i TLDs.
- 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**
- Ott. 20 – Nov. 20 **Incarico di collaborazione professionale: “Sviluppo di nuove funzionalità di un software basato su intelligenza artificiale per la riabilitazione domiciliare”**  
Camlin Italy Srl
- Ho sviluppato nuove funzionalità di un software basato su intelligenza artificiale per la riabilitazione

domiciliare, in particolare ho disegnato e modificato il modello dei dati relazionale (PostgreSQL) per l'inserimento di una classe di esercizi terapeutici e creato i relativi endpoint REST utilizzati dalla app Android dedicata. Ho modificato l'integrazione della libreria di machine learning per la valutazione della nuova classe di esercizi. Ho integrato un endpoint websocket per la comunicazione realtime dei flussi di dati tra app e backend.

**Attività o settore** Progettazione e sviluppo software

**Ago. 20 – Set. 20** **Incarico di collaborazione professionale: “Sviluppo di un recommendation system”**  
MeeTravel srl

MeeTravel ha sviluppato un'applicazione che permette a degli utenti di partecipare a viaggi organizzati attraverso la app. Il mio ruolo è stato quello di sviluppare un recommendation system attraverso tecniche di machine learning (clustering e classificatore supervisionato) per definire la compatibilità tra un utente ed uno specifico viaggio, compresi i partecipanti ad esso.

**Attività o settore** Ricerca e Sviluppo

**Ott. 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

In allegato, in coda al cv, la certificazione fornita dal Dipartimento di Informatica dell'Università di Pisa con protocollo 0001542/2022 del 27/07/2022.

Progettazione e sviluppo di un sistema per il rilevamento di anomalie sul traffico di rete attraverso l'utilizzo di Deep Learning; il sistema, basato su LSTM, è stato allenato su un dataset annotato contenente attacchi di rete, come ad esempio attacchi DOS, Brute Force, Sql Injection ecc ed ha ottenuto ottime performance per il riconoscimento degli attacchi di rete.

Progettazione e sviluppo di un parser a dipendenze basato su Deep Learning e scritto in PyTorch, con il quale ha permesso la partecipazione alla competizione internazionale “IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies”. 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 e scritto in PyTorch. 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** **Incarico di collaborazione professionale: Software Architect / Data Scientist**  
Dataniinja 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 Dataniinja 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 *PySpark*.

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, che sfrutta l'albero di parsing e il part of speech del testo per effettuare le correzioni. Il sistema attualmente genera quotidianamente podcast su dati finanziari e sulla situazione Covid.

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

## Curriculum vitae

Attività o settore Ricerca e Sviluppo

Mar. 18–Mar. 19 **Dipendente fulltime a tempo indeterminato: 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

In allegato, in coda al cv, la certificazione fornita dal Dipartimento di Informatica dell'Università di Pisa.

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, durante la collaborazione con l'Istituto di Informatica e Telematica del CNR 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-sartiano/DeepWebClassifier>).

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

È 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 internazionale 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 dei server del gruppo, basate su sistemi Linux, tra cui una macchina avente 4 GPU Tesla P100 per effettuare calcoli ad alte prestazioni, utili per creare modelli di Deep Learning.

Attività o settore Ricerca e Sviluppo

Feb. 12–Feb. 18 **Co-fondatore e incarico di collaborazione professionale**

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

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,

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.

Ho progettato e sviluppato parte del prodotto principale di BioCare (drDrin), un sistema per mettere in comunicazione 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.

Ero responsabile della parte sperimentale sull'analisi di referti medici attraverso tecniche di NLP e machine learning. Sono stato tutor aziendale per i tirocinanti di Informatica e Informatica umanistica dell'Università di Pisa, i quali hanno lavorato alla progettazione di alcuni verticali del software aziendale.

**Attività o settore** e-health

**Set. 16–Gen. 17** **Incarico di collaborazione professionale: Analisi automatica di traffico di rete criptato**

ntop di Luca Deri

Disegno e implementazione di uno strumento per che analizza il traffico di rete criptato e riconosce specifiche sequenze di traffico, come ad esempio l'invio di messaggi su Twitter. Lo strumento è implementato come un classificatore supervised basato su una rete neurale profonda (Deep Learning), ed è composto principalmente da layer LSTM. Il classificatore accetta in input sequenze di pacchetti di rete, in cui per ogni pacchetto sono state estratte delle feature dal software ntop. L'implementazione è stata effettuata in Python utilizzando il framework Tensorflow per l'implementazione del classificatore e Redis per l'invocazione di esso in maniera asincrona.

**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)

In allegato, in coda al cv, la certificazione fornita dal Dipartimento di Informatica dell'Università di Pisa.

Partecipazione al progetto di ricerca POR CReO FESR TOSCANA 2007-2013 “RIS Ricerca e Innovazione in ambito socio sanitario”, dove sono stati progettati e realizzati strumenti in grado di estrarre informazioni di tipo medico da referti clinici espressi in linguaggio naturale in modo automatico. Durante il progetto è stato inoltre creato un corpus di apprendimento utilizzabile per allenare un modello di Named Entity Recognition.

Gli strumenti fanno parte di una completa pipeline linguistica, e utilizzano tecniche di machine learning per l'apprendimento automatico di tipo supervisionato e non supervisionato. La pipeline è scritta in C++, e tramite delle API create con il framework SWIG, sono stati forniti dei binding per poterla invocare in altri in altri linguaggi, come Python, Java e PHP.

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)

In allegato, in coda al cv, la certificazione fornita dal Dipartimento di Informatica dell'Università di Pisa  
Creazioni di corpus paralleli, alcuni ottenuti direttamente dal web estraendo contenuti dalle pagine html.



## Curriculum vitae

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 C++ messe a disposizione da Moses, creando inoltre la parte di frontend in HTML/Javascript. Il deploy del sistema è stato effettuato su macchine linux.

Sono stato anche manutentore della pipeline linguistica del gruppo di Human Language Technologies del Dipartimento di Informatica, pipeline scritta in C++, che tramite delle API, create con il framework SWIG, ha permesso di fornire dei binding per poterla invocare in altri in altri linguaggi, come Python, Java e PHP.

**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*, attraverso 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.tornadoweb.org/> ).
- Definizione delle feature e successiva fase di features selection. Le *feature* sono state selezionate 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 **Stage: Sviluppatore Software**

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

In questa Società ho iniziato apprendendo le nozioni fondamentali nell'area CTI (Computer Telephony Integration), 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. Per poter procedere alla stima del rischio, sono state individuate le vulnerabilità della rete tramite il software Nessus, il quale ci ha permesso di modellarli attraverso un attack-tree, i cui nodi corrispondono a degli attacchi che sfruttano le vulnerabilità trovate. Successivamente è stato costruito un grafo pesato e orientato delle dipendenze tra le vulnerabilità identificate, in cui vulnerabilità dipendenti sono connesse tramite archi pesati. È stato poi applicato un algoritmo iterativo che aggiorna i pesi delle vulnerabilità in base alle proprie dipendenze con altre vulnerabilità. In ultimo, tramite i pesi aggiornati agli step precedenti, è stato aggiornato il rischio associato alle minacce in maniera ricorsiva sull'attack-tree, identificando gli attacchi più fattibili e di conseguenza le vulnerabilità da sanare in prima istanza.

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	Letture	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:**

Amministrazione di ambienti 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, Matlab.

### CONTAINER:

Docker, Docker Compose.

### SVILUPPO MOBILE:

Android, Ionic 2.

### SVILUPPO WEB:

Django, Flask, FastAPI, Tornado, 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 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.

### Sistema di prevenzione di infortuni sportivi basato su Intelligenza Artificiale

Ho progettato e sviluppato un sistema che analizza dati sportivi e li mette in relazione agli infortuni, in modo tale da poterlo utilizzare per prevenire gli infortuni sportivi.

Il prototipo del sistema è stato costruito e testato utilizzando i dati della squadra Zebre Rugby Club di Parma.

Tra i dati disponibili vi erano: infortuni, dati GPS utilizzato durante gli allenamenti, dati del nutrizionista e dei preparatori atletici.

Ho creato una pipeline di analisi dati, in cui tutti i dati disponibili sono preprocessati e normalizzati e per ogni membro della squadra è stata creata una serie temporale con i relativi dati; di queste serie temporali sono state poi create delle finestre temporali a cui sono state associate le informazioni relative ad un eventuale infortunio. Questi sono state utilizzate come input al sistema di

apprendimento automatico in fase di allenamento.

Il sistema è stato scritto in Python e sono state utilizzate prevalentemente le librerie *pandas* per il preprocessing dei dati, *sklearn* per la feature selection e il classificatore supervisionato basato su XGBoost è stato implementato tramite la libreria *xgboost*.

Patente di guida A, B

## ULTERIORI INFORMAZIONI

### Attività formativa

- Dal 2013 fornisco supporto al docente del corso di *Natural Language Processing/Human Language Technologies*, Professor Giuseppe Attardi del Dipartimento di Informatica dell'Università di Pisa, nel seguire e aiutare gli studenti per lo sviluppo dei progetti d'esame e per le loro tesi di laurea triennale e magistrale. I temi toccati dagli studenti comprendono lo stato dell'arte nelle tecniche di Natural Language Processing, tra cui l'utilizzo di metodologie basate su deep e machine learning.
- 25 Marzo 2019 – "Il pensiero computazionale – Tides Google" – Ho tenuto un seminario per professori di scuole superiori, insegnando i principi del linguaggio di programmazione *Python* e illustrando un esempio di utilizzo del *Reinforcement Learning*.

### Produzione Scientifica

- Crema C, Attardi G, Sartiano D, Redolfi A. Natural language processing in clinical neuroscience and psychiatry: A review. *Front Psychiatry*. 2022 Sep 14;13:946387. doi: 10.3389/fpsy.2022.946387. PMID: 36186874; PMCID: PMC9515453.
- D. Sartiano, A. Del Soldato. "READS: REgistrant Anomalies Detection System - An Anomaly Detection System of the .it Registry database", Technical Report "IIT TR-05/2022 " (CNR-IIT), 2022.
- 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 clinical notes." *CLiC it (2015)*: 14.
- Luca Deri and Maurizio Martinelli and Daniele Sartiano and Loredana Sideri. "Large Scale Web-Content Classification". *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 2015.
- Attardi, Giuseppe, Vittoria Cozza, and Daniele Sartiano. "Adapting Linguistic Tools for the Analysis

## Curriculum vitae

of Italian Medical Records." *CliC-it*. 2014.

- Attardi, Giuseppe, Vittoria Cozza, and Daniele Sartiano. "UniPi: Recognition of mentions of disorders in clinical text." *SemEval 2014* (2014): 754.
- Attardi Giuseppe, Andrea Buzzelli, and Daniele Sartiano. "Machine Translation for Entity Recognition across Languages in Biomedical Documents." (2013).
- Raffaello Brondi, Filippo Bannò, Sara Bendinelli, Christian Castelli, Antonio Mancina, Mauro Marinoni, Daniele Sartiano, Francesca Semissi. "Drugs Don't Work in Patients Who Don't Take Them: Dr. Drin, The New ICT Paradigm For Chronic Therapies." *Annual Review of Cybertherapy and Telemedicine 2013: Positive Technology and Health Engagement for Healthy Living and Active Ageing* 191 (2013): 100.
- Attardi, Giuseppe, Daniele Sartiano, and Maria Simi. "Active Learning for Domain Adaptation of Dependency Parsing on Legal Texts." *Semantic Processing of Legal Texts (SPLeT-2012) Workshop Programme*. 2012.

### Trattamento dei dati personali

Autorizzo il trattamento dei dati personali contenuti nel mio curriculum vitae in base all'art. 13 del D. Lgs. 196/2003 e all'art. 13 del Regolamento UE 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali.