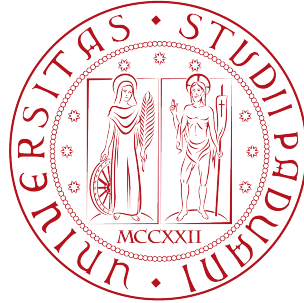


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"



Scuola di Scienze

CORSO DI LAUREA IN INFORMATICA

## *Piano di Lavoro*

STUDENTE Valentina Signor  
MATRICOLA 1049106

AZIENDA Zucchetti S.P.A

data

## Contatti

**Studente:** Valentina Signor, [valentina.signor@studenti.unipd.it](mailto:valentina.signor@studenti.unipd.it), + 39 342 083 14 26;

**Tutor aziendale:** Gregorio Piccoli, [gregorio.piccoli@zucchetti.it](mailto:gregorio.piccoli@zucchetti.it), + 39 0371 59 457 11;

**Azienda:** Zucchetti S.P.A, Via Giovanni Cittadella 7, 735137 - Padova, <https://www.zucchetti.it>. soluzioni software, hardware e servizi.

## Scopo dello stage

*Zucchetti S.P.A.*, con la sua esperienza di oltre 40 anni rappresenta la prima software house italiana per storia e dimensione. Fondata nel 1978 ad opera di Domenico Zucchetti, per commercializzare in tutta Italia un software che garantisca la gestione automatizzata della dichiarazione dei redditi, ad oggi non solo permette ad Aziende, Professionisti, Associazioni di Categoria e Pubblica Amministrazione di disporre di soluzioni software, hardware e servizi innovativi, ma si è diffusa anche all'estero presentando sedi in Francia, Germania, Romania, Spagna, Svizzera, Brasile e Stati Uniti.

Nell'innovativa offerta che *Zucchetti S.P.A.* mette a disposizione dei propri clienti si possono identificare una vasta gamma di prodotti operanti negli ambiti:

- gestionali: contabilità, acquisti, vendite, magazzino ecc.;
- delle Risorse umane: paghe, stipendi, presenze, controllo accessi ecc.;
- fiscali: dichiarazione dei redditi, 730, conservazione sostitutiva ecc.

Oltre che molti altri prodotti di contorno nei campi della Business Intelligence, robotica, IoT e sicurezza.

Lo stage prevede l'inserimento dello studente nella realtà aziendale al fine di garantirne la formazione attraverso compiti reali e sperimentazione diretta dei processi e contesti lavorativi.

Il progetto da sviluppare permetterà di ricavare a partire dall'osservazione di fenomeni reali spiegati solo attraverso *Leggi empiriche*, delle leggi e teorie quantificabili e concrete. Esse permetteranno calate nell'ambito degli algoritmi di *Machine Learning*, di fornire delle solide basi ai dati osservati in modo da identificare un pattern comportamentale che li spieghi.

Lo studente dovrà seguire i seguenti passi:

- osservazione di *Leggi empiriche* e di Teorie non formali;
- descrizione delle varie fasi di analisi del dato; essa si articolerà attraverso preparazione e pulizia del dato, analisi preliminare, esposizione dei dati, analisi predittiva e analisi prescrittiva;
- analisi dei dati nell'ambito del *Supervised Learning* mediante adattamento e applicazione di algoritmi di regressione (Regressione Lineare) e di classificazione (SVM e Random Forest);
- analisi dei dati nell'ambito del *Unsupervised Learning* attraverso adattamento e applicazione del k-Means Clustering;
- applicazione delle Reti Bayesiane al fine di stabilire delle leggi;
- riepilogo critico e documentato per ciascuna delle due categorie di apprendimento, dei risultati ottenuti.

## Interazione tra studente e tutor aziendale

Vista la situazione particolare in cui verte il nostro Paese da inizio marzo 2020 (causa COVID-19), si assicura che lo stage in caso di necessità possa essere svolto anche in modalità smart-working. In tal caso le regole a cui attenersi per verificare lo stato di avanzamento del lavoro, chiarirne eventualmente gli obiettivi e aggiornare il Piano di Lavoro se necessario, saranno le seguenti:

- lo studente si terrà comunque in stretta comunicazione col proprio tutor aziendale Gregorio Piccoli; ciò dovrà avvenire almeno quotidianamente, attraverso attività di video-conferenza o tramite altri canali digitali;
- sarà necessario un riconoscimento dell'impegno orario dallo studente; quest'ultimo provvederà quindi a redigere un registro, informale ma preciso, che giornalmente verrà sottoposto a convalida dal tutor aziendale. Sarà proprio tale registro ad abilitare la firma del tutor aziendale sul modulo di fine stage, dando testimonianza dello svolgimento delle ore di lavoro richieste.

Nel caso in cui lo stage potesse essere svolto in presenza lo studente farà sempre riferimento al tutor aziendale Gregorio Piccoli con il quale lavorerà a stretto contatto.

Tale strategia si auspica di facilitare l'integrazione dello studente all'interno dell'ambiente lavorativo adeguandolo alle caratteristiche del periodo corrente.

## **Contenuti formativi previsti**

Durante questo progetto di stage lo studente avrà occasione di approfondire le sue conoscenze:

- nell'osservazione e analisi critica dei dati attraverso:
  - osservazione di fenomeni reali al fine di ricavarne regole precise che ne spieghino il coportamento;
  - l'uso del software Orange Canvas;
  - tecniche di pulizia del dato.
- nelle aree del *Machine Learning* attraverso:
  - lo studio, l'adattamento e l'applicazione di differenti algoritmi di analisi e predizione sui dati;
  - utilizzo di tecniche di Supervised ed Unsupervised Learning.

## Pianificazione del Lavoro

### Pianificazione settimanale

- **Prima Settimana - Studio del problema attraverso Osservazioni, Leggi empiriche e Teorie**

- preparazione e analisi dei dati;
- studio del software Orange Canvas.

Durante questa settimana lo studente impiegherà parte del monte delle ore previste ad inquadrare il problema, predisponendo il lavoro che si articolerà nelle settimane successive. Si approccerà al software Orange Canvas studiandone il funzionamento, facendone uso per facilitare almeno inizialmente l'attività di analisi dei dati. In particolare si dedicherà alla preparazione e alla pulizia dei valori osservati, per poi passare alla loro esportazione e analisi (preliminare, predittiva e prescrittiva). Potrà inoltre osservare i comportamenti dei vari algoritmi di Machine Learning confrontandone le performance e la precisione. Lo studente sarà quindi chiamato ad effettuare uno studio approfondito circa le modalità di osservazione e manipolazione dei dati in modo da localizzarne le conseguenti regole empiriche caratterizzanti.

- **Seconda Settimana - Programma per la Regressione Lineare in Javascript e Java**

- studio dell'algoritmo di Regressione Lineare;
- studio e adattamento della libreria corrispondente;
- aggiunta previsione in Java.

Durante questa settimana lo studente si impegnerà a prendere confidenza con l'algoritmo di Regressione Lineare, anche grazie all'utilizzo di Orange Canvas; successivamente procederà all'adattamento della relativa libreria Javascript fornita dall'azienda: l'obiettivo è quello di renderla omogenea con le altre librerie "sorelle" relative agli altri algoritmi di Machine Learning. Lo studente cercherà quindi di creare per quanto possibile, un'interfaccia comune mettendo mano dove abbia senso, alla struttura, ai metodi, alle variabili della libreria in modo da ottenere un risultato che sia il più manutenibile possibile. Provvederà poi adottando il linguaggio Java, ad estendere tale libreria aggiungendovi la possibilità di fare previsioni. Tale funzionalità non dovrebbe dimostrarsi particolarmente macchinosa in quanto consisterebbe semplicemente dell'aggiunta di alcune linee di codice differenti in base alla classe di algoritmo considerata.

- **Terza Settimana - Programma per la Support Vector Machines in Javascript e Java**

- studio dell'algoritmo di Support Vector Machine;
- studio e adattamento della libreria corrispondente;
- aggiunta previsione in Java.

Durante questa settimana lo studente si impegnerà a prendere confidenza con l'algoritmo di Support Vector Machine sempre aiutandosi con Orange Canvas; successivamente procederà come trattato al punto sopra all'adattamento della relativa libreria Javascript ed alla sua estensione attraverso l'aggiunta del algoritmo di previsione.

- **Quarta Settimana - Test e Documentazione**

Durante questa settimana, lo stagista documenterà e testerà la bontà del codice da lui prodotto durante le settimane precedenti.

- **Quinta Settimana - Programma per k-Means in Javascript e Java**

- studio dell'algoritmo non supervisionato k-Means;
- studio e adattamento della libreria corrispondente;

- aggiunta previsione in Java.

Durante questa settimana lo studente svolgerà uno studio individuale sugli algoritmi non supervisionati ponendo particolare attenzione al k-Mean. Cercherà di interfacciarsi grazie anche all'ausilio del software Orange Canvas procedendo come fatto anche per gli algoritmi supervisionati, all'adattamento della relativa libreria Javascript ed alla sua estensione attraverso l'aggiunta del algoritmo di previsione.

- **Sesta Settimana - Programma per Random Forest in Javascript e Java**

- studio delle Random Forest;
- studio e adattamento della libreria corrispondente;
- aggiunta previsione in Java.

Durante questa settimana lo studente passerà allo studio delle Random Forest, ne esaminerà le caratteristiche e il comportamento grazie anche all'uso di Orange Canvas; successivamente procederà come fatto per gli altri algoritmi, all'adattamento della relativa libreria Javascript fornita dall'azienda ed alla sua estensione attraverso l'aggiunta del algoritmo di previsione.

- **Settima Settimana - Programma per Reti Bayesiane in Javascript e Java**

- studio delle Reti Bayesiane;
- studio e adattamento della libreria corrispondente;
- aggiunta previsione in Java.

Durante questa settimana lo studente si impegnerà a prendere confidenza con le Reti Bayesiane, studiandone le caratteristiche e il comportamento grazie anche all'uso di Orange Canvas. Con esse lo studente potrà localizzare delle teorie che gli permetteranno di spiegare attraverso una base ragionata i dati osservati inizialmente. Successivamente procederà come fatto per gli altri algoritmi all'adattamento della relativa libreria Javascript fornita dall'azienda, ed alla sua estensione attraverso l'aggiunta del algoritmo di previsione.

- **Ottava Settimana - Test e Documentazione**

Durante questa settimana, lo stagista come fatto anche dopo le prime tre settimane di lavoro, provvederà a documentare e testare la bontà del codice da lui prodotto durante le settimane precedenti.

## Ripartizione oraria

La pianificazione, in termini di quantità di ore di lavoro, verrà distribuita come presentato di seguito.

Durata in ore	Data (inizio - fine)	Descrizione attività
40	2020.06.08 - 2020.06.12	Studio del problema attraverso Osservazioni, Leggi empiriche e Teorie.
40	2020.06.15 - 2020.06.19	Programma per la Regressione Lineare in Javascript e Java.
40	2020.06.22 - 2020.06.26	Programma per la Support Vector Machines in Javascript e Java.
40	2020.06.29 - 2020.06.03	Test e Documentazione.
40	2020.07.06 - 2020.07.10	Programma per k-Means in Javascript e Java.
40	2020.07.13 - 2020.07.17	Programma per Random Forest in Javascript e Java.
40	2020.07.20 - 2020.07.24	Programma per Reti Bayesiane in Javascript e Java.
40	2020.07.27 - 2020.07.31	Test e Documentazione.

## Diagramma di Gantt

Di seguito viene riportato il diagramma di Gantt relativo al Piano di Lavoro previsto.

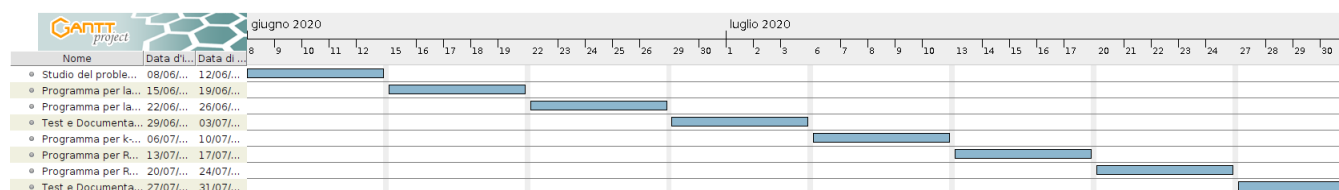


Figura 1: Piano di Lavoro suddiviso per le 8 settimane previste dallo stage

## Obiettivi

### Notazione

Si farà riferimento ai requisiti seguendo le seguenti notazioni:

- **O**: utilizzato per i *requisiti obbligatori*, sono da considerarsi vincolanti in quanto obiettivo primario richiesto dal committente;
- **D**: utilizzato per i *requisiti desiderabili*, sono da considerarsi non vincolanti o strettamente necessari, ma dal riconoscibile valore aggiunto;
- **F**: utilizzato per i *requisiti facoltativi*, sono da considerarsi portatori di valore aggiunto non strettamente competitivo.

Le sigle precedentemente indicate saranno seguite da una coppia sequenziale di numeri, rappresentati l'identificativo del requisito.

Obbligatori	
O01	Individuazione e analisi di Leggi empiriche e Teorie.
O02	Studio degli algoritmi di Regressione Lineare e Support Vector Machine.
O03	Studio degli algoritmi di k-Means, Random Forest e Reti Bayesiane.
O04	Creazione di un'interfaccia comune per gli algoritmi di Machine Learning analizzati.
O05	Implementazione previsione in Java.
O06	Produzione di una documentazione completa come resoconto delle attività svolte e delle osservazioni effettuate.
O07	Mantenimento della riservatezza necessaria nei confronti dei materiali forniti e dei dati osservati durante lo svolgimento dello stage.

Desiderabili	
D01	Per ogni algoritmo la previsione in SQL.

Facoltativi	
F01	Reti Bayesiane + storia documentazione.

## Prodotti attesi a fine stage

Di seguito vengono esposti i prodotti che si attende dallo stage:

Obbligatoria

Desiderabili

Facoltativi