Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

CORSO DI LAUREA IN INFORMATICA



Utilizzo dei Modelli di Machine Learning di AWS per la Classificazione e l'Estrapolazione di Informazioni contenute nelle Mail PEC

Tesi di Laurea Triennale

Relatore

Prof. Lamberto Ballan

Laure and o

Riccardo Zaupa

Matricola 2034303





Ringraziamenti

Desidero esprimere la mia gratitudine al professor Lamberto Ballan, mio relatore, per l'aiuto e il sostegno che mi ha dato durante la stesura dell'elaborato. Vorrei anche ringraziare, con affetto, i miei genitori per il loro sostegno, il grande aiuto e la loro presenza in ogni momento durante gli anni di studio. Desidero poi ringraziare i miei amici per i bellissimi anni trascorsi insieme e le mille avventure vissute.

Padova, Settembre 2024

Riccardo Zaupa

Sommario

Il presente documento descrive il lavoro svolto durante il periodo di stage svolta presso l'azienda Sanmarco Informatica S.p.A.. Lo stage, svolto alla conclusione del percorso di studi triennale in Informatica presso l'Università degli Studi di Padova, ha avuto una durata complessiva di trecentoore. L'obiettivo principale dello stage è stato quello di classificare e estrapolare informazioni contenute nelle mail PEC utilizzando i modelli di Machine Learning di AWS.

Indice

1	Inti	roduzione	1
	1.1	L'azienda	2
	1.2	L'idea	2
	1.3	Organizzazione del testo	3
2	Pro	ocessi e metodologie	4
	2.1	Processo sviluppo prodotto	4
3	Des	scrizione dello stage	5
	3.1	Introduzione al progetto	5
	3.2	Analisi preventiva dei rischi	6
	3.3	Requisiti e obiettivi	6
	3.4	Pianificazione	7
		3.4.1 Pianificazione settimanale	7
4	Ana	alisi dei requisiti	9
	4.1	Casi d'uso	9
	4.2	Requisiti	10
	4.3	Tracciamento dei requisiti	11
5	Pro	ogettazione e codifica	12
	5.1	Tecnologie e strumenti	12
	5.2	Ciclo di vita del software	12
	5.3	Progettazione	12
		5.3.1 Namespace 1	12
	5.4	Design Pattern utilizzati	12
	5.5	Codifica	12
6	Ver	rifica e validazione	13
7	Cor	nclusioni	14
	7.1	Consuntivo finale	14

INDICE

7.2	Raggiungimento degli obiettivi	14
7.3	Conoscenze acquisite	14
7.4	Valutazione personale	14
Acroni	mi e abbreviazioni	15
Glossai	rio	16
Bibliog	grafia	17

Elenco delle figure

1.1	Lorem	1
2.1	Lorem	4
3.1	Caption	5
4.1	Use Case - UC0: Scenario principale	9
6.1	Lorem	13

Elenco delle tabelle

4.1	Tabella del tracciamento dei requisti funzionali	10
4.2	Tabella del tracciamento dei requisiti qualitativi	1
4.3	Tabella del tracciamento dei requisiti di vincolo	1

Introduzione



Figura 1.1: Lorem

Introduzione al contesto applicativo.

Lorem Figure 1.1

Esempio di utilizzo di un termine nel glossario Application Program Interface (API).

Esempio di citazione in linea site:agile-manifesto.

Esempio di citazione nel pie' di pagina citazione¹

Termine di glossario API text

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit

¹womak:lean-thinking.

mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

1.1 L'azienda

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.2 L'idea

Introduzione all'idea dello stage². Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada portitior diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel

²article:spooky.

magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

1.3 Organizzazione del testo

Il secondo capitolo descrive ...

Il terzo capitolo approfondisce ...

Il quarto capitolo approfondisce ...

Il quinto capitolo approfondisce ...

Il sesto capitolo approfondisce ...

Nel settimo capitolo descrive ...

Riguardo la stesura del testo, relativamente al documento sono state adottate le seguenti convenzioni tipografiche:

- gli acronimi, le abbreviazioni e i termini ambigui o di uso non comune menzionati vengono definiti nel glossario, situato alla fine del presente documento;
- i termini in lingua straniera o facenti parti del gergo tecnico sono evidenziati con il carattere corsivo.

Processi e metodologie

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. article:spooky

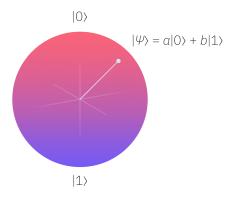


Figura 2.1: Lorem

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.1 Processo sviluppo prodotto

Descrizione dello stage

3.1 Introduzione al progetto



Figura 3.1: Caption

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.2 Analisi preventiva dei rischi

Durante la fase di analisi iniziale sono stati individuati alcuni possibili rischi a cui si potrà andare incontro. Si è quindi proceduto a elaborare delle possibili soluzioni per far fronte a tali rischi.

1. Performance del simulatore hardware

Descrizione: le performance del simulatore hardware e la comunicazione con questo potrebbero risultare lenti o non abbastanza buoni da causare il fallimento dei test.

Soluzione: coinvolgimento del responsabile a capo del progetto relativo il simulatore hardware.

3.3 Requisiti e obiettivi

Gli obiettivi sono stati definiti in accordo con il tutor aziendale e si identificano nel seguente modo:

[Priorità][Id]

- Priorità: indica la priorità dell'obiettivo, può essere obbligatorio o desiderabile;
- Id: identifica l'obiettivo in modo univoco rispetto alla priorità.

ID	Categoria	Descrizione		
O01	Obbligatorio	Analisi dei servizi AWS per l'addestramento dei modelli AI		
O02	Obbligatorio	Addestramento di un modello di apprendimento AI utilizzando i		
		servizi AWS		
O03	Obbligatorio	Analisi requisiti applicativi e tecnici per implementare la soluzione		
		richiesta		
O04	Obbligatorio	Implementare un modello di apprendimento automatico che anal-		
		izzi il contenuto delle ^[g] PEC importate e assegni loro categorie		
		appropriate in base al contenuto (mittente, destinatario, data e		
		argomento)		
D01	Desiderabile	Implementare algoritmi di IA in grado di adattarsi e appren-		
		dere continuamente dai dati per migliorare le prestazioni del sis-		
		tema nel tempo. Ciò include l'ottimizzazione dei modelli di ap-		
		prendimento automatico in base all'esperienza e ai feedback degli		
		utenti		
D02	Desiderabile	Integrazione con un sistema documentale per l'archiviazione delle		
		PEC creando i metadati necessari con le informazioni estratte e		
		collocandole nella corretta categoria di appartenenza		

3.4 Pianificazione

3.4.1 Pianificazione settimanale

Il periodo di stage è stato suddiviso in 8 settimane, durante le quali sono previste le seguenti attività:

Settimana	Dal	Al	Attività	
1	24-06-2024	28-06-2024	- Incontro con persone coinvolte nel progetto per dis-	
			cutere i requisiti e le richieste di implementazione	
			- Ricerca, studio e documentazione per inquadra-	
			mento progetto	
			- Introduzione ai linguaggi di sviluppo	
			- Introduzione agli ambienti di sviluppo	
			- Introduzione dei servizi ^[g] awsg	
2	01-07-2024	05-07-2024	- Analisi dei servizi AWS per l'addestramento di un	
			modello di apprendimento	
			- Addestramento di un modello di apprendimento	
			utilizzando i servizi di AWS	
			Milestone: Utilizzo dei servizi AWS per l'addestra-	
			mento di un modello di apprendimento	
3	08-07-2024	12-07-2024	- Studio della soluzione per definire i requisiti neces-	
			sari per l'implementazione	
			Milestone: Analisi dei requisiti applicativi e tecnici	
			per implementare la soluzione	
4	15-07-2024	19-07-2024	- Addestramento modello di apprendimento per	
			catalogare le PEC in base al loro contenuto	
5	22-07-2024	26-07-2024	- Implementazioni per interfacciarsi con il modello di	
			apprendimento addestrato e per poter catalogare le	
			PEC importate	
			Milestone: Completamento obiettivi minimi	
6	29-07-2024	02-08-2024	- Implementazione algoritmo di AI per l'autoap-	
			prendimento	
7	05-08-2024	09-08-2024	- Studio e documentazione sulle API messe a dispo-	
			sizione dal documentale per poter catalogare le mail	
			PEC	
			- Implementazione dell'integrazione con il documen-	
			tale producendo i metadati necessari per catalogare	
			le PEC	

CAPITOLO 3. DESCRIZIONE DELLO STAGE

Settimana	Dal	Al	Attività	
8	12-08-2024	16-08-2024	- Verifica e test archiviazione PEC nel documentale	
		Milestone: Completamento obiettivi massimi		
9	19-08-2024	23-08-2024	4 - Recupero eventuali ritardi	
10	26-08-2024	30-08-2024	- Recupero eventuali ritardi	

Analisi dei requisiti

4.1 Casi d'uso

Per lo studio dei casi di utilizzo del prodotto sono stati creati dei diagrammi. I diagrammi dei casi d'uso (in inglese *Use Case Diagram*) sono diagrammi di tipo Unified Modeling Language (UML) dedicati alla descrizione delle funzioni o servizi offerti da un sistema, così come sono percepiti e utilizzati dagli attori che interagiscono col sistema stesso. Essendo il progetto finalizzato alla creazione di un tool per l'automazione di un processo, le interazioni da parte dell'utilizzatore devono essere ovviamente ridotte allo stretto necessario. Per questo motivo i diagrammi d'uso risultano semplici e in numero ridotto.

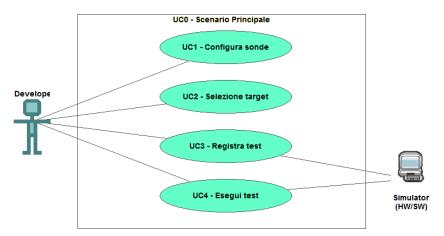


Figura 4.1: Use Case - UC0: Scenario principale

UC0: Scenario principale

Attori Principali: Sviluppatore applicativi.

Precondizioni: Lo sviluppatore è entrato nel plug-in di simulazione all'interno dell'IDE.

Descrizione: La finestra di simulazione mette a disposizione i comandi per configurare, registrare o eseguire un test.

Postcondizioni: Il sistema è pronto per permettere una nuova interazione.

UC1: Gestione Utente

Attori Principali: Amministratore, Utente Registrato.

Precondizioni: L'utente deve essere autenticato nel sistema..

Descrizione: L'utente può gestire le informazioni del proprio profilo..

Postcondizioni: Le modifiche vengono salvate nel sistema..

Scenario Alternativo: Se l'utente non è autenticato, visualizza un messaggio di errore..

UC2: Creazione Prodotto

Attori Principali: Amministratore.

Precondizioni: L'amministratore ha effettuato l'accesso al sistema..

Descrizione: L'amministratore può aggiungere un nuovo prodotto al catalogo...

Postcondizioni: Il nuovo prodotto viene aggiunto con successo...

Scenario Alternativo: Se i campi obbligatori non sono compilati, visualizza un messaggio di errore..

4.2 Requisiti

Da un'attenta analisi dei requisiti e degli use case effettuata sul progetto è stata stilata la tabella che traccia i requisiti in rapporto agli use case.

Sono stati individuati diversi tipi di requisiti e si è quindi fatto utilizzo di un codice identificativo per distinguerli.

Il codice dei requisiti, dove ogni requisito è identificato con il carattere \mathbf{R} , è così strutturato:

F: Funzionale.

Q: Qualitativo.

V: Di vincolo.

N: Obbligatorio (necessario).

D: Desiderabile.

Z: Opzionale.

Nelle tabelle 4.1, 4.2 e 4.3 sono riassunti i requisiti e il loro tracciamento con gli use case delineati in fase di analisi.

Requisito	Descrizione	Use Case
RFN-1	L'interfaccia permette di configurare il tipo di sonde del test	UC1

Tabella 4.1: Tabella del tracciamento dei requisti funzionali

CAPITOLO 4. ANALISI DEI REQUISITI

Requisito	Descrizione	Use Case
RQD-1	Le prestazioni del simulatore hardware deve garantire la giusta esecuzione	-
	dei test e non la generazione di falsi negativi	

Tabella 4.2: Tabella del tracciamento dei requisiti qualitativi

Requisito	Descrizione	Use Case
RVO-1	La libreria per l'esecuzione dei test automatici deve essere riutilizzabile	-

Tabella 4.3: Tabella del tracciamento dei requisiti di vincolo

4.3 Tracciamento dei requisiti

Progettazione e codifica

Breve introduzione al capitolo

5.1 Tecnologie e strumenti

Di seguito viene data una panoramica delle tecnologie e strumenti utilizzati.

Tecnologia 1

Descrizione Tecnologia 1.

Tecnologia 2

Descrizione Tecnologia $2\,$

5.2 Ciclo di vita del software

5.3 Progettazione

5.3.1 Namespace 1

Descrizione namespace 1.

Classe 1: Descrizione classe 1

Classe 2: Descrizione classe 2

5.4 Design Pattern utilizzati

5.5 Codifica

Verifica e validazione



Figura 6.1: Lorem

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Conclusioni

7.1 Consuntivo finale

Ipsum

7.2 Raggiungimento degli obiettivi

Sit amet

- 7.3 Conoscenze acquisite
- 7.4 Valutazione personale

Acronimi e abbreviazioni

```
API Application Programming Interface. i, 1, 14, 16
AWS Amazon Web Services. i, 16
PEC Posta Elettronica Certificata. i, 16
SDK Software Development Kit. i, 14, 16
UML Unified Modeling Language. i, 9, 16
```

Glossario

- API text In informatica con il termine API si indica ogni insieme di procedure disponibili al programmatore, di solito raggruppate a formare un set di strumenti specifici per l'espletamento di un determinato compito all'interno di un certo programma. La finalità è ottenere un'astrazione, di solito tra l'hardware e il programmatore o tra software a basso e quello ad alto livello semplificando così il lavoro di programmazione. i, 1, 14, 15
- AWS Amazon Web Services (AWS) è una piattaforma di servizi cloud che offre potenza di calcolo, storage di database, distribuzione di contenuti e altre funzionalità per aiutare le imprese a scalare e crescere. i, 15
- **PEC** La *Posta Elettronica Certificata* (PEC) è un servizio di posta elettronica che garantisce l'invio e la ricezione di messaggi di posta elettronica con valore legale equivalente a quello della raccomandata con avviso di ricevimento.. i, 15
- SDK text A software development kit (SDK) is a collection of software development tools in one installable package. They facilitate the creation of applications by having a compiler, debugger and sometimes a software framework. They are normally specific to a hardware platform and operating system combination. To create applications with advanced functionalities such as advertisements, push notifications, etc; most application software developers use specific software development kits.
 i, 14, 15
- UML text In ingegneria del software Unified Modeling Language (ing. linguaggio di modellazione unificato) è un linguaggio di modellazione e specifica basato sul paradigma object-oriented. L'UML svolge un'importantissima funzione di "lingua franca" nella comunità della progettazione e programmazione a oggetti. Gran parte della letteratura di settore usa tale linguaggio per descrivere soluzioni analitiche e progettuali in modo sintetico e comprensibile a un vasto pubblico. i, 15

Bibliografia