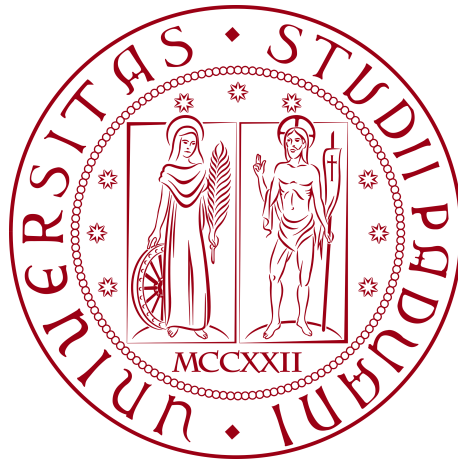


**Università degli Studi di Padova**

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

*Laureando*

Riccardo Zaupa

Matricola 2034303



*“I’m stronger, I’m smarter, I’m better”*

— Homelander.

*“Diventerò il re dei pirati”*

— Monkey D. Luffy.

*“Non devo fuggire”*

— Shinji Ikari.

*“Non dire gatto se non ce l’hai nel sacco”*

— Nonna.

*“GG\MM\AAAAAAA”*

— Alex Scantamburlo.

*“Ucciderò tutti i giganti”*

— Eren Yeager.

*“Con il superamento della revisione PB – a fronte di una prestazione estremamente deludente – avete concluso il vostro progetto didattico di IS.”*

— Tullio Vardanega.

*“Non posso aiutarvi”*

— Alessandro Staffolani.

*“Mi avete fatto perdere mesi della mia vita”*

— Riccardo Cardin.

# 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

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	L'azienda . . . . .	1
1.2	L'offerta di stage . . . . .	2
1.3	Organizzazione del testo . . . . .	2
<b>2</b>	<b>Processi e metodologie</b>	<b>3</b>
2.1	Processo sviluppo prodotto . . . . .	3
<b>3</b>	<b>Descrizione dello stage</b>	<b>4</b>
3.1	Introduzione al progetto . . . . .	4
3.2	Analisi preventiva dei rischi . . . . .	5
3.3	Requisiti e obiettivi . . . . .	5
3.4	Pianificazione . . . . .	6
3.4.1	Pianificazione settimanale . . . . .	6
<b>4</b>	<b>Analisi dei requisiti</b>	<b>8</b>
4.1	Casi d'uso . . . . .	8
4.2	Requisiti . . . . .	9
4.3	Tracciamento dei requisiti . . . . .	10
<b>5</b>	<b>Progettazione e codifica</b>	<b>11</b>
5.1	Tecnologie e strumenti . . . . .	11
5.2	Ciclo di vita del software . . . . .	11
5.3	Progettazione . . . . .	11
5.3.1	Namespace 1 . . . . .	11
5.4	Design Pattern utilizzati . . . . .	11
5.5	Codifica . . . . .	11
<b>6</b>	<b>Verifica e validazione</b>	<b>12</b>
<b>7</b>	<b>Conclusioni</b>	<b>13</b>
7.1	Consuntivo finale . . . . .	13

7.2	Raggiungimento degli obiettivi . . . . .	13
7.3	Conoscenze acquisite . . . . .	13
7.4	Valutazione personale . . . . .	13
<b>Acronimi e abbreviazioni</b>		<b>14</b>
<b>Glossario</b>		<b>15</b>
<b>Bibliografia</b>		<b>16</b>

# Elenco delle figure

1.1	Logo di Sanmarco Informatica . . . . .	1
2.1	Lorem . . . . .	3
3.1	Caption . . . . .	4
4.1	Use Case - UC0: Scenario principale . . . . .	8
6.1	Lorem . . . . .	12

# Elenco delle tabelle

4.1	Tabella del tracciamento dei requisiti funzionali . . . . .	9
4.2	Tabella del tracciamento dei requisiti qualitativi . . . . .	10
4.3	Tabella del tracciamento dei requisiti di vincolo . . . . .	10

# Capitolo 1

## Introduzione

### 1.1 L'azienda

Sanmarco Informatica S.p.A. è un'azienda italiana di sviluppo software e consulenza informatica. Da quarant'anni supporta la riorganizzazione dei processi in tutte le aree aziendali, progettando e realizzando soluzioni digitali integrate.

L'azienda che ad oggi conta più di 600 dipendenti e oltre 2500 aziende seguite ha come sede principale Villa Ramanelli a Grisignano di Zocco, in provincia di Vicenza, poco distante dai Centri di Ricerca e Sviluppo (CRS) e dal Centro per la Formazione di Vicenza. Conta anche diverse filiali in Trentino-Alto Adige, Friuli-Venezia Giulia, Lombardia, Piemonte, Emilia-Romagna, Toscana, Campania e Puglia.

L'azienda è organizzata in Business Unit, dei centri di competenza specifici e autonomi ma in relazione costante. Ognuna delle quali è specializzata in un settore specifico.

L'obiettivo principale è l'innovazione e il progresso tecnologico, con l'obiettivo di creare soluzioni software che siano in grado di rispondere alle esigenze dei clienti, garantendo la massima qualità e sicurezza.

La metodologia di lavoro, indipendentemente dalla Business Unit, è basata su un approccio Agile implementata con il framework Scrum. Agile è un approccio alla gestione dei progetti che si basa su principi di collaborazione, auto-organizzazione e flessibilità. Scrum è un framework Agile che permette di gestire progetti complessi, garantendo la massima trasparenza e la massima flessibilità.

Ulteriori informazioni sono disponibili sul sito web dell'azienda<sup>1</sup>.



**Figura 1.1:** Logo di Sanmarco Informatica

---

<sup>1</sup><https://www.sanmarcoinformatica.com/>



## 1.2 L'offerta di stage

L'idea dello stage consiste nella catalogazione delle Poste Elettroniche Certificate (PEC) integrando tecnologie di Intelligenza Artificiale (AI) per l'analisi e l'efficienza del processo.

Il modello di apprendimento automatico analizza il contenuto delle PEC e le classifica in base al contenuto.

Il progetto è stato proposto dall'azienda in occasione dell'evento Stage IT 2024, organizzato dall'Università degli Studi di Padova e promosso da Confindustria Veneto Est. Tale evento mira ad agevolare l'incontro tra studenti e aziende, offrendo la possibilità di svolgere uno stage formativo con specifico riferimento al settore ICT.

## 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 ...

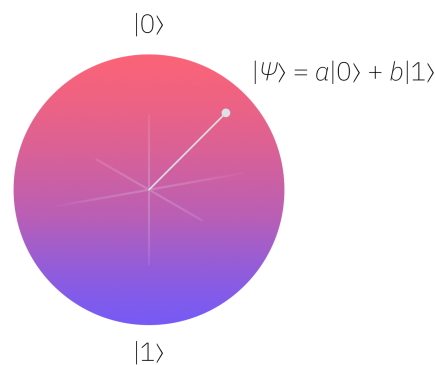
Riguardo la stesura del testo, relativamente al documento sono state adottate le seguenti convenzioni tipografiche:

- 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*.

## Capitolo 2

# Processi e metodologie

Lorem ipsum dolor sit amet, consectetur 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, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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

## Capitolo 3

# Descrizione dello stage

### 3.1 Introduzione al progetto



**Figura 3.1:** Caption

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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: composto da due cifre, 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 analizzi il contenuto delle <sup>[g]</sup> 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 apprendere continuamente dai dati per migliorare le prestazioni del sistema nel tempo. Ciò include l'ottimizzazione dei modelli di apprendimento 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	<ul style="list-style-type: none"><li>- Incontro con persone coinvolte nel progetto per discutere i requisiti e le richieste di implementazione</li><li>- Ricerca, studio e documentazione per inquadramento progetto</li><li>- Introduzione ai linguaggi di sviluppo</li><li>- Introduzione agli ambienti di sviluppo</li><li>- Introduzione dei servizi <sup>[g]</sup>aws</li></ul>
2	01-07-2024	05-07-2024	<ul style="list-style-type: none"><li>- Analisi dei servizi AWS per l'addestramento di un modello di apprendimento</li><li>- Addestramento di un modello di apprendimento utilizzando i servizi di AWS</li></ul> <b>Milestone:</b> Utilizzo dei servizi AWS per l'addestramento di un modello di apprendimento
3	08-07-2024	12-07-2024	<ul style="list-style-type: none"><li>- Studio della soluzione per definire i requisiti necessari per l'implementazione</li></ul> <b>Milestone:</b> Analisi dei requisiti applicativi e tecnici per implementare la soluzione
4	15-07-2024	19-07-2024	<ul style="list-style-type: none"><li>- Addestramento modello di apprendimento per catalogare le PEC in base al loro contenuto</li></ul>
5	22-07-2024	26-07-2024	<ul style="list-style-type: none"><li>- Implementazioni per interfacciarsi con il modello di apprendimento addestrato e per poter catalogare le PEC importate</li></ul> <b>Milestone:</b> Completamento obiettivi minimi
6	29-07-2024	02-08-2024	<ul style="list-style-type: none"><li>- Implementazione algoritmo di AI per l'autoapprendimento</li></ul>
7	05-08-2024	09-08-2024	<ul style="list-style-type: none"><li>- Studio e documentazione sulle API messe a disposizione dal documentale per poter catalogare le mail PEC</li><li>- Implementazione dell'integrazione con il documentale producendo i metadati necessari per catalogare le PEC</li></ul>

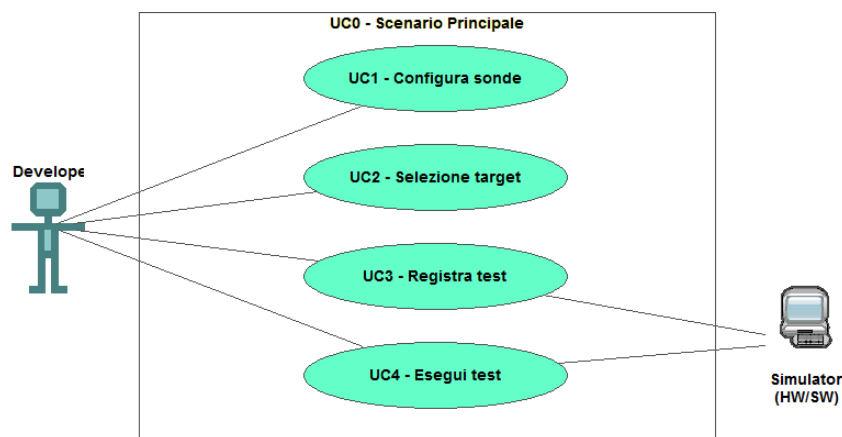
Settimana	Dal	Al	Attività
8	12-08-2024	16-08-2024	- Verifica e test archiviazione PEC nel documentale <b>Milestone:</b> Completamento obiettivi massimi
9	19-08-2024	23-08-2024	- Recupero eventuali ritardi
10	26-08-2024	30-08-2024	- Recupero eventuali ritardi

## Capitolo 4

# 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 **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 requisiti funzionali



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

# Capitolo 5

## 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

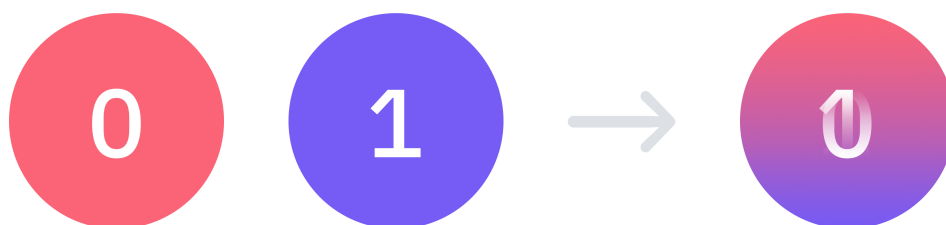
**Classe 2:** Descrizione classe 2

### 5.4 Design Pattern utilizzati

### 5.5 Codifica

## Capitolo 6

### Verifica e validazione



**Figura 6.1:** Lorem

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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.

## Capitolo 7

# Conclusioni

Lorem <sup>[g]</sup>[SDK text](#)

Lorem [API text](#)

### 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, 13, 15

**AWS** [Amazon Web Services](#). i, 15

**PEC** [Posta Elettronica Certificata](#). i, 15

**SDK** [Software Development Kit](#). i, 13, 15

**UML** [Unified Modeling Language](#). i, 8, 15

# 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](#), [13](#), [14](#)

**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](#), [14](#)

**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](#), [14](#)

**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](#), [13](#), [14](#)

**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](#), [14](#)

# Bibliografia