Immagine che contiene decorato, colorato, guarnizioni

Descrizione generata automaticamente

**Autori: Simone Della Porta, Antonio De Lucia, Antonio Maddaloni, Rocco Iuliano, Francesco Peluso**

1. **Business Understanding**
   1. **Introduzione al problema**
   2. **Obiettivi**
   3. **Descrizione dell’ambiente**
      1. **Specifica PEAS**
   4. **Business Success Criteria**
   5. **Tool da utilizzare**
2. **Data Understanding**
   * 1. **Scelta del dataset**
     2. **Analisi del dataset**
3. **Data Preparation**
   * 1. **Data Cleaning**
     2. **Feature Scaling**
     3. **Feature Engineering**
     4. **Data Balancing**
4. **Data Modeling**
5. **Evaluation**
6. **Deployment**
7. **Business Understanding**
   1. **Introduzione al problema**

Con l’avvenire della pandemia del covid-19 iniziata nel 2019 molte organizzazioni sanitarie stanno cercando di tener traccia dei contagi e del tasso di mortalità dovuti ad essa; tutto ciò per limitare l’aumento della curva dei contagi.

* 1. **Obiettivi**

Dato il problema in esame, l’obiettivo del nostro progetto è realizzare un modello di machine learning che sia capace di predire se una persona risulta essere positiva o negativa al covid-19 sulla base di alcuni sintomi.

* 1. **Descrizione dell’ambiente**
     1. **Specifica PEAS**

|  |  |
| --- | --- |
| **PEAS** | |
| **Performance** | La misura di performance dell’agente è la sua capacità di avvicinarsi quanto più possibile ad una situazione ideale nella quale vengono predetti correttamente i veri positivi e negativi al covid-19. |
| **Environment** | L’ambiente in cui opera il nostro agente riguarda l’ambito sanitario e più in particolare la virologia. L’ambiente è:   * **Statico** in quanto l’ambiente non cambia nel tempo e non cambia alle azioni effettuate dell’agente. * **Episodico** in quanto un’azione intrapresa dall’agente in un dato istante non è influenzata dall’azione effettuata precedentemente. * **Completamente osservabile** in quanto l’agente in ogni istante può avere una visione completa dell’ambiente in cui è calato. * **Discreto** in quanto l’agente può effettuare soltanto determinate azioni e ricevere determinati impulsi. * **Non noto** in quanto l’agente non conosce il risultato delle proprie azioni. * **Singolo** in quanto l’ambiente prevede al proprio interno l’introduzione di un singolo agente.   Gli elementi dell’ambiente sono le persone che hanno determinati sintomi e che possono essere possibili positivi al covid-19. |
| **Actuators** | L’agente agisce sull’ambiente tramite lo stream di output del nostro computer nel quale andrà a predire la positività di una persona. |
| **Sensors** | L’agente riceverà gli impulsi tramite lo stream di input del nostro computer. |

* 1. **Business success criteria**

Il criterio con il quale andremo a validare il nostro sistema è il seguente:

* + - * + Il modello deve avere un buon livello di accuracy, ovvero almeno il 50%. Questo perché essendo un problema di natura medica e trattiamo un virus nuovo, vuol dire che i dati a disposizione per far apprendere il modello non sono sicuramente accurati. Inoltre è da notare che il modello predice la positività o meno di una persona in base ai sintomi e non è detto che se una persona ha determinati sintomi allora ha covid ma potrebbe avere altre patologie che comportano sintomi simili. Un ulteriore motivo per il quale abbiamo scelto questa soglia di accuracy è che il modello potrà sbagliare molto facilmente con persone asintomatiche in quanto non hanno sintomi e risultano come persone “negative” agli “occhi” del modello.
  1. **Tool da utilizzare**

I tool da utilizzare per sviluppare il progetto sono:

* + - * + Visual Studio Code
        + Python
        + Pandas
        + GitHub
        + Scikitlearn
        + Mathplot
        + Kaggle
        + Overleaf

1. **Data Understanding**
   1. **Scelta del dataset**
   2. **Analisi del dataset**
2. **Data Preparation**
   1. **Data Cleaning**
   2. **Feature Scaling**
   3. **Feature Engineering**
   4. **Data Balancing**
3. **Data Modeling**
4. **Evaluation**
5. **Deployment**