## Documentazione del Codice

### Descrizione Generale:

Il codice fornito è un’applicazione web implementata in Flask che permette agli utenti di caricare un file Excel contenente dati da etichettare, visualizzare i dati nel browser e successivamente esportare i risultati etichettati in un nuovo file Excel.

### Dipendenze:

Il codice utilizza le seguenti librerie Python: - Flask - pandas - random - re - scikit-learn

### Struttura del Codice:

Il codice è diviso in tre parti principali: il backend Flask, il frontend HTML e il modulo Model definito in module.py.

#### Flask Backend:

Il backend Flask gestisce il routing delle richieste HTTP e la logica di business dell’applicazione. È costituito da tre rotte principali:

1. **/** (index): Questa rotta restituisce una pagina HTML per caricare il file Excel.
2. **/upload**: Questa rotta gestisce il caricamento del file Excel, il suo processamento e la visualizzazione dei dati nel browser.
3. **/export**: Questa rotta gestisce l’esportazione dei dati etichettati in un nuovo file Excel.

#### Frontend HTML:

Il frontend è costituito da due file HTML:

1. **index.html**: Questo file contiene un semplice form per il caricamento del file Excel.
2. **label.html**: Questo file visualizza i dati del file Excel caricato e consente agli utenti di etichettare i dati.

#### Modulo Model:

Il modulo Model contiene la logica per preparare i dati, addestrare il modello RandomForest e effettuare le predizioni. Le funzioni principali sono:

* **prepare\_data**: Prepara i dati per l’addestramento del modello, rimuovendo colonne non necessarie, eseguendo la TF-IDF su testo preprocessato e combinando i dati numerici con le caratteristiche estratte dal testo.
* **train\_randomforest**: Addestra un modello RandomForest sui dati preparati.
* **predict**: Effettua una predizione su un singolo dato.
* **get\_size**: Restituisce la dimensione del DataFrame.

### Descrizione delle Funzionalità:

1. **Caricamento del File Excel**: Gli utenti possono caricare un file Excel contenente dati da etichettare.
2. **Visualizzazione dei Dati**: I dati vengono visualizzati nel browser per permettere agli utenti di etichettarli.
3. **Etichettatura dei Dati**: Gli utenti possono etichettare i dati selezionando una delle opzioni disponibili per ciascun dato.
4. **Esportazione dei Dati Etichettati**: Una volta etichettati, i dati possono essere esportati in un nuovo file Excel.

### Utilizzo dell’Applicazione:

Per utilizzare l’applicazione, l’utente deve seguire questi passaggi:

1. Accedere all’URL dell’applicazione nel browser.
2. Caricare un file Excel contenente i dati da etichettare.
3. Visualizzare i dati e etichettarli secondo le opzioni disponibili.
4. Esportare i dati etichettati in un nuovo file Excel.

### Note di Implementazione:

- Il modello RandomForest viene addestrato utilizzando i dati presenti nel file Excel caricato.  
- Il testo dei dati viene preprocessato utilizzando la tecnica TF-IDF prima di essere utilizzato per l'addestramento del modello.  
- L'etichettatura dei dati avviene selezionando una delle opzioni disponibili per ciascun dato e poi esportando i risultati in un nuovo file Excel.