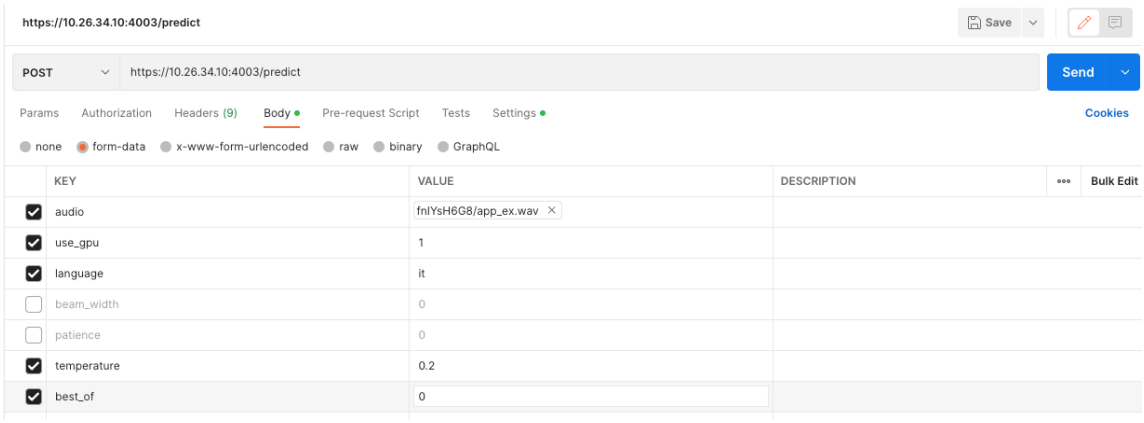
ASR-IT: WHISPER

1. Utilizzo del servizio tramite client web

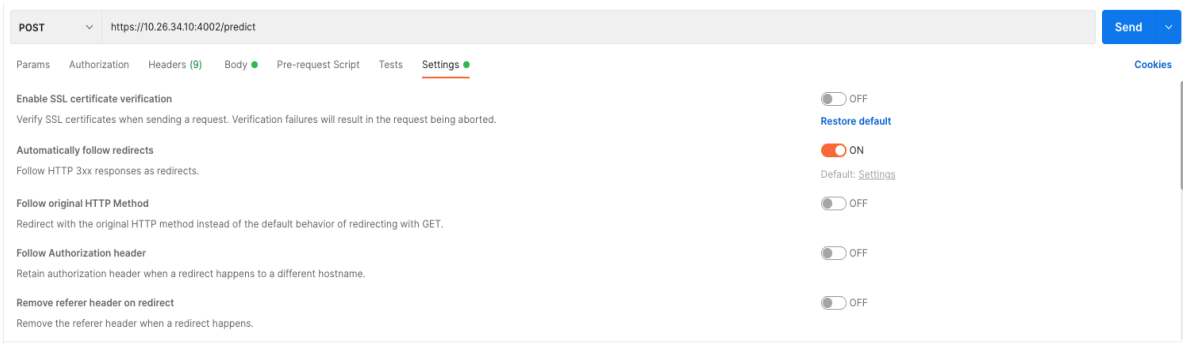
1. Connettersi a https://10.26.34.10:4003/ e accettare i rischi dovuti alla mancanza di certificati.
2. Selezionare un file da caricare o registrare tramite microfono. La qualità del microfono potrebbe influenzare l'accuratezza della trascrizione.
3. Selezionare i parametri scelti per l' inferenza. Per maggiori informazioni sulle opzioni vedere sezione 3: Parametri.

2. Utilizzo del servizio tramite Postman

1. Scaricare e creare un account Postman.
2. Nella sezione per effettuare richieste HTTP, inserire l'indirizzo https://10.26.34.10:4003/predict e selezionare *form-data*.
3. Selezionare **body ➤ form data**, per maggiori informazioni sulle opzioni vedere sezione 3: Parametri.



4. Disattivare la verifica SSL nella sezione impostazioni.



5. Effettuare la richiesta ed attendere la risposta (vedere sezione 4: Risposta).

3. Parametri

I parametri selezionabili sono i seguenti:

- **use_gpu**: se usare la GPU in inferenza. *Valori ammessi: [0,1]. Default: 0.*

  ⚠️ Se si richiede l'utilizzo della GPU, il modello dovrà essere spostato da CPU a GPU, prima dell' inferenza. Lo spostamento richiede del tempo (monitorabile dal campo 'processing times' della risposta).

  ⚠️ La GPU potrebbe non essere libera o non essere abbastanza potente per supportare il modello.

- **language:** linguaggio del file caricato. È possibile rilevare automaticamente il linguaggio con l'opzione 'DETECT'. Altrimenti è necessario fornire una stringa che rappresenta il linguaggio: 'it' per italiano, 'en' per inglese etc etc. Vedere sezione 5: Linguaggi per le opzioni.

Il decoding può avvenire o in modo greedy, con sampling o con beam search (più accurato ma più lento).

A. Per eseguire il decoding in modo greedy non è necessario specificare parametri.
B. I parametri da specificare per beam search sono:

   - **beam_width**: ampiezza delle opzioni considerate in fase di inferenza con language model. Default: 0

   - **patience**: patience del beam search (Vedere paper). Default 0.

C. I parametri da specificare per sampling decoding sono:

   - **temperature**: Di quanto aumentare la probabilità dei token più probabili. Default: 0. Range: [0,1].

   - **best_of**: Numero di samples indipendenti da campionare, se t > 0. Default: 0.

⚠️ I parametri **beam_width** e **patience** riguardano la beam search, mentre **temperature** e **best_of** riguardano il sampling decoding. Se si specificano opzioni per **beam search** o **patience** non ha senso specificarne anche per **temperature** e **best_of**.

4. Risposta

La risposta contiene i seguenti campi:

- **processing_times**: elenco dei tempi di esecuzione di ogni fase dell' inferenza.

- **settings**: parametri della richiesta.

- **language:** linguaggio (su richiesta rilevato automaticamente).

- **results**: trascrizione.
- i**nfo**: informazioni addizionali su stato server, se necessarie.

---

5. Linguaggi

L'accuracy del modello multilingual non è la stessa per tutti i linguaggi (vd tabella).
Linguaggi ammessi per Whisper multilingual:

```
"en": "english",
"zh": "chinese",
"de": "german",
"es": "spanish",
"ru": "russian",
"ko": "korean",
"fr": "french",
"ja": "japanese",
"pt": "portuguese",
"tr": "turkish",
"pl": "polish",
"ca": "catalan",
"nl": "dutch",
"ar": "arabic",
"sv": "swedish",
"it": "italian",
"id": "indonesian",
"hi": "hindi",
"fi": "finnish",
"vi": "vietnamese",
"iw": "hebrew",
"uk": "ukrainian",
"el": "greek",
"ms": "malay",
"cs": "czech",
"ro": "romanian",
"da": "danish",
"hu": "hungarian",
"ta": "tamil",
"no": "norwegian",
"th": "thai",
"ur": "urdu",
"hr": "croatian",
"bg": "bulgarian",
"lt": "lithuanian",
"la": "latin",
"mi": "maori",
"ml": "malayalam",
"cy": "welsh",
"sk": "slovak",
"te": "telugu",
"fa": "persian",
"lv": "latvian",
"bn": "bengali",
"sr": "serbian",
"az": "azerbaijani",
"sl": "slovenian",
"kn": "kannada",
"et": "estonian",
"mk": "macedonian",
"br": "breton",
"eu": "basque",
"is": "icelandic",
"hy": "armenian",
"ne": "nepali",
"mn": "mongolian",
"bs": "bosnian",
"kk": "kazakh",
"sq": "albanian",
"sw": "swahili",
"gl": "galician",
"mr": "marathi",
"pa": "punjabi",
"si": "sinhala",
"km": "khmer",
"sn": "shona",
"yo": "yoruba",
"so": "somali",
"af": "afrikaans",
"oc": "occitan",
"ka": "georgian",
"be": "belarusian",
"tg": "tajik",
"sd": "sindhi",
"gu": "gujarati",
"am": "amharic",
"yi": "yiddish",
"lo": "lao",
"uz": "uzbek",
"fo": "faroese",
```

```
"ht": "haitian creole",
"ps": "pashto",
"tk": "turkmen",
"nn": "nynorsk",
"mt": "maltese",
"sa": "sanskrit",
"lb": "luxembourgish",
"my": "myanmar",
"bo": "tibetan",
"tl": "tagalog",
"mg": "malagasy",
"as": "assamese",
"tt": "tatar",
"haw": "hawaiian",
"ln": "lingala",
"ha": "hausa",
"ba": "bashkir",
"jw": "javanese",
"su": "sundanese",
```