**LANGUAGE DETECTION WITH NLP**

Download Dataset [here](https://drive.google.com/file/d/1SM5soYEOAKbvv0GczXWkKlebfgsNwbWX/view?usp=share_link)

### **About the Dataset**

It's a small language detection dataset. This dataset consists of text details for 17 different languages.

**Languages**

1) English  
2) Malayalam  
3) Hindi  
4) Tamil  
5) Kannada  
6) French  
7) Spanish  
8) Portuguese  
9) Italian  
10) Russian  
11) Swedish  
12) Dutch  
13) Arabic  
14) Turkish  
15) German  
16) Danish  
17) Greek

**TASKS**

1. Import the dataset and perform pre processing on the data.
2. Clean the data of unnecessary symbols.
3. Train the models on the data after performing a train test split.
4. Evaluate the models and plot their results
5. Write a function that will take a sentence in one of the above 17 languages and will predict what language is the sentence written in.  
   Predict on the following test cases:
   1. **Case 1**: "Innomatics Research labs provides a community based knowledge portal for full stack and Data Science professionals"
   2. **Case 2**: "Innomatics Research labs fournit un portail de connaissances basé sur la communauté pour les professionnels de l'analyse et de la science des données"
   3. **Case 3**: "توفر Innomatics Research labs بوابة معرفية قائمة على المجتمع لمحترفي التحليلات وعلوم البيانات"
   4. **Case 4**: "Innomatics Research labs proporciona un portal de conocimiento basado en la comunidad para profesionales de Analytics y Data Science."
   5. **Case 5**: “അനലിറ്റിക്സ്, ഡാറ്റാ സയൻസ് പ്രൊഫഷണലുകൾക്കായി കമ്മ്യൂണിറ്റി അധിഷ്ഠിത വിജ്ഞാന പോർട്ടൽ അനലിറ്റിക്സ് വിദ്യ നൽകുന്നു"
   6. **Case 6**: "Innomatics Research labs - это портал знаний на базе сообщества для профессионалов в области аналитики и данных."