**Project Name:** Bharatlaw Translation

**Problem Statement:**

Translate a given text summary into Gujarati language using python.

**Objective:**

1. Research different translation technologies like Google Translate, Hugging Face LLMs and implement all researched technologies to translate the given summary.
2. Translation should be contextual and concise enough to understand easily, instead of just word to word translation.
3. Present findings and recommend best technology for further development.

**Approach:**

* At first, I read and understood the objective and task of the given project. Then I searched about available translation technologies using Bard/Gemini, ChatGPT, google search and listed down them for implementation.
* Then I researched about each model and learned how to implement them.
* After that, I implemented the models individually and tested their result.
* Lastly, I choose the best models based on their results and performances.

**Software used:**

Google Collab, Kaggle notebook

**Result:**

* **Models I Found out:**

1. Google Translate
2. Google Madlad-400 (Hugging Face LLM)
3. Bloom
4. Facebook/mbart-large-50-many-to-many-mmt (Hugging Face LLM)
5. Facebook/m2m100\_418M (Hugging Face LLM)

* **Google Translate performance**:
* It gave very good and pretty accurate result.
* Sample translation of the given summary in Gujarati:

વાટાઘાટોવાળા સાધનો અધિનિયમની કલમ 138 હેઠળ ફરિયાદ અંગેનો ચુકાદો

રજૂઆત

આ લખાણમાં વાટાઘાટયોગ્ય ઇન્સ્ટ્રુમેન્ટ્સ એક્ટની કલમ 138 હેઠળ નોંધાયેલી ફરિયાદ અંગે ભારતની સુપ્રીમ કોર્ટના ચુકાદાની ચર્ચા કરવામાં આવી છે.આ કેસમાં પ્રતિવાદી દ્વારા જારી કરાયેલા ચેક અંગેના વિવાદનો સમાવેશ થાય છે, જે અપૂરતા ભંડોળને કારણે પરત કરવામાં આવ્યું હતું.ટ્રાયલ કોર્ટે શરૂઆતમાં ફરિયાદને ફગાવી દીધી હતી, પરંતુ સુપ્રીમ કોર્ટે તેને સમર્થન આપ્યું હતું, અને તે શોધી કા .્યું હતું કે આ ચેક ખરેખર જવાબ આપનાર દ્વારા જારી કરવામાં આવ્યો હતો.

કી પોઇન્ટ

Apple પલ કાર્ટનની સંખ્યા અને બાકી રકમ અંગેના પુરાવાઓના વિરોધાભાસને કારણે શરૂઆતમાં ફરિયાદ રદ કરવામાં આવી હતી.

હાઈકોર્ટે સ્થાપિત કર્યું છે કે જ્યાં સુધી અન્યથા સાબિત ન થાય ત્યાં સુધી ચેક વિચારણાની ધારણા ધરાવે છે.

પુરાવા અથવા સંજોગો આપીને કોઈ દેવું અસ્તિત્વમાં નથી તે અંગેના પુરાવા અથવા સંજોગો આપીને વિચારણાની ધારણાને નકારી કા to વા માટે પુરાવાનો ભાર છે.

અદાલતે કાયદાની કલમ ૧ 139 139 હેઠળ દેવું અથવા જવાબદારીની ધારણા અંગે ચર્ચા કરી છે અને જણાવે છે કે જો આરોપી સંભવિત સંરક્ષણ વધારે છે તો તે નિષ્ફળ થઈ શકે છે.

કોર્ટ ભારપૂર્વક જણાવે છે કે કલમ ૧ 139 139 હેઠળની ધારણા એ મુકદ્દમામાં અયોગ્ય વિલંબને રોકવા માટેનું એક ઉપકરણ છે અને ચેકનું અપમાન કરવું એ મોટાભાગે નાગરિક ખોટું છે.

આ કેસમાં જવાબ આપનાર ચેક જારી કરવામાં વિચારણાની ધારણાને નકારી કા to વા માટે કોઈ પુરાવા પૂરા પાડવામાં નિષ્ફળ ગયો.

નીચેની અદાલતોની રકમના નિર્ધારમાં વિસંગતતાઓના આધારે ફરિયાદને નકારી કા .વા બદલ ટીકા કરવામાં આવી હતી.

જવાબ આપનારને ચેકનું અપમાન કરવા બદલ દોષી ઠેરવવામાં આવે છે અને તેને દંડ અને ખર્ચ ચૂકવવાનો આદેશ આપવામાં આવે છે.

અંત

નિષ્કર્ષમાં, ભારતની સુપ્રીમ કોર્ટે વાટાઘાટોવાળા ઇન્સ્ટ્રુમેન્ટ્સ એક્ટની કલમ 138 હેઠળ નોંધાયેલી ફરિયાદને સમર્થન આપ્યું હતું.કોર્ટે શોધી કા .્યું હતું કે જવાબ આપનાર દ્વારા ચેક જારી કરવામાં આવ્યો હતો અને પુરાવામાં વિસંગતતાના આધારે ફરિયાદને નકારી કા .વા માટે નીચલી અદાલતોની ટીકા કરી હતી.કોર્ટે કલમ ૧ 139 139 હેઠળ વિચારણાની ધારણા પર ભાર મૂક્યો હતો અને જવાબ આપનારને ચેકનું અપમાન કરવા બદલ દોષી ઠેરવ્યો હતો.જવાબ આપનારને દંડ અને ખર્ચ ચૂકવવાનો આદેશ આપવામાં આવ્યો હતો.

* Google translate is very fast at translating large texts. It generates results in 2-5 seconds.
* **Google Madlad-400 performance**:
* It also gave very good and pretty decent result.
* Sample translation example of given summary in Gujrati:
* <2gu> my name is surajit. i am from west bengal. i am eating an apple.

મારું નામ સુરજિત છે. હું પશ્ચિમ બંગાળનો છું. હું એક સફરજન ખાઈ રહ્યો છું.

* <2gu> In conclusion, the Supreme Court of India upheld a complaint filed under Section 138 of the Negotiable Instruments Act. The court found that the cheque was issued by the respondent and criticized the lower courts for dismissing the complaint based on discrepancies in the evidence. The court emphasized the presumption of consideration under Section 139 and held the respondent guilty of dishonoring the cheque. The respondent was ordered to pay a fine and costs.

અંતે, ભારતની સુપ્રીમ કોર્ટે નેટવર્કિંગ ઇન્સ્ટ્રુમેન્ટ્સ એક્ટની કલમ 138 હેઠળ દાખલ કરવામાં આવેલી ફરિયાદને માન્ય રાખી હતી. કોર્ટે શોધી કાઢ્યું હતું કે, ચેક જવાબદારે જારી કર્યો હતો અને સાક્ષીઓમાં વિરોધાભાસના આધારે ફરિયાદને ફગાવી દેવા બદલ નીચલી અદાલતોની ટીકા કરી હતી. કોર્ટે કલમ 139 હેઠળની ગણતરી પર ભાર મૂક્યો હતો અને જવાબદારને ચેકનું અપમાન કરવા બદલ ગુનેગાર ઠેરવ્યો હતો. જવાબદારને દંડ અને ખર્ચ ચૂકવવાનો આદેશ આપ્યો હતો.

* Drawbacks:

It doesn’t give perfect results every time.

It lacks in performance, if we provide large texts, in a single time. It’s best to send 4-5 lines of text at a time. It shall help in quick response and more accurate response.

* **Bloom:**
* It can translate in Gujarati, but doesn’t give accurate results often. One of the reason is that the model is trained on only 0.04% Gujarati language and therefore needs finetuning.

Sample example:  
Translate the following sentences from english to gujarati.

Text: my name is surajit. i am from west bengal. i am eating an apple.

Translated text: નામ: સરોજિત. હું પશ્ચિમ બંગાળનો રહેવાસી છું. હું એક પ્લમ ખાઉં છું.</s>

Translate the following sentences from english to gujarati.

Text: In conclusion, the Supreme Court of India upheld a complaint filed under Section 138 of the Negotiable Instruments Act. The court found that the cheque was issued by the respondent and criticized the lower courts for dismissing the complaint based on discrepancies in the evidence. The court emphasized the presumption of consideration under Section 139 and held the respondent guilty of dishonoring the cheque. The respondent was ordered to pay a fine and costs.

Translated text: 结论માં, સુપ્રીમ કોર્ટ ઓફ ઇન્ડિયાએ એક ફરિયાદ દાખલ કરી હતી કે તે કલમ 138 હેઠળ નાણા દસ્તાવેજો કાયદાના હેઠળ દાખલ કરવામાં આવી હતી. કોર્ટે તે ચુકાદો આપ્યો કે ચેક રિટર્ન કરનાર છે અને તે ફરિયાદને નકારવામાં આવી હતી કારણ કે દસ્તાવેજો સાથે ભિન્ન પુરાવા હતા. કોર્ટે પ્રારૂપની માન્યતા કાયદા હેઠળ 139 હેઠળ અને આરોપ મૂક્યો હતો કે ચેક રિટર્ન કરનાર છે અને તેને નકારવામાં આવ્યો હતો કારણ કે દસ્તાવેજો સાથે ભિન્ન પુરાવા હતા. કોર્ટે પ્રારૂપની માન્યતા કાયદા હેઠળ 139 હેઠળ અને આરોપ મૂક્યો હતો કે ચેક રિટર્ન કરનાર છે અને તેને નકારવામાં આવ્યો હતો કારણ કે દસ્તાવેજો સાથે ભિન્ન પુરાવા હતા. કોર્ટે પ્રારૂપની માન્યતા કાયદા હેઠળ 139 હેઠળ અને આરોપ મૂક્યો હતો કે ચેક રિટર્ન કરનાર છે અને તેને નકારવામાં આવ્યો હતો કારણ કે દસ્તાવેજો સાથે ભિન્ન પુરાવા હતા. કોર્ટે પ્રારૂપની માન્યતા કાયદા હેઠળ 139 હેઠળ અને આરોપ મૂક્યો હતો કે ચેક રિટર્ન કરનાર છે અને તેને નકારવામાં આવ્યો હતો કારણ કે દસ્તાવેજો સાથે ભિન્ન પુરાવા હતા. કોર્ટે પ્રારૂપની માન્યતા કાયદા હેઠળ 139 હેઠળ અને આરોપ મૂક્યો હતો કે ચેક રિટર્ન કરનાર છે અને તેને નકારવામાં આવ્યો હતો કારણ કે દસ્તાવેજો સાથે ભિન્ન પુરાવા હતા.</s>

**Future Scope:**

As most of the LLM models are trained on general purpose use, text generation and question answering. They are not very good at translating.  
But if we train those pre-trained models with datasets, we can get better results.  
here is a related research paper: <https://arxiv.org/pdf/2311.09216.pdf>  
here is an available dataset for Gujarati language : <https://ai4bharat.iitm.ac.in/bpcc/>

**Resources used:**

I mostly followed hugging face documentations to learn about the models and YouTube tutorials for reference.

here are few resources: https://huggingface.co/docs/transformers/tasks/translation, https://arxiv.org/pdf/2311.09216.pdf, https://huggingface.co/facebook/mbart-large-50-many-to-many-mmt, https://huggingface.co/google/madlad400-3b-mt, https://pypi.org/project/googletrans/ etc.