AIML Project Client Report

An Interpreter to Predict Language

2320030244 – R. Jayaram

2320030361 – S. Shanmukha

2320030294 – P. Abhiram

Client-1:

1. What languages do you want the system to recognize and predict?
We want the system to predict around 20 major languages, including English, Spanish, French, Chinese, Arabic, Hindi, etc.

2. What input types will users provide — text, audio, or both?
Initially, users will input text only. In the future, we may expand to audio input.

3. Should the model predict only the language, or also translate or transcribe?
For now, the system should only predict the language. No translation or transcription is needed at this stage.

4. What level of accuracy are you expecting for language prediction?
We are targeting at least 90% prediction accuracy on clean and properly formatted input data.

5. Do you have existing datasets for training the model, or should we find/create one?
We don’t have a dataset yet. You will need to find a public dataset or build a custom one for this project.

6. Should the interpreter work in real-time, or is slight delay acceptable?
A slight delay is acceptable, as long as it does not exceed 2–3 seconds after input is given.

7. Where will the interpreter be deployed — mobile app, web app, desktop app, or embedded device?
We are planning to deploy it as a web application accessible via browsers on both PC and mobile devices.

8. Do you require additional features like identifying regional dialects (e.g., British vs. American English)?
No, identifying the main language is enough. We don't need classification now.

9. How should the system handle unknown languages?

If the input language is unknown, the system should return a message like "Language not recognized" instead of making random guesses.

10. What privacy or security standards must be followed when handling user data?

User data must be kept private. No storing of input data on servers unless it's strictly necessary for model training and with user consent.

Client-2:

1. Will the system need to support multiple inputs at once (batch predictions)?

Yes, it should support both single and batch inputs for faster processing.

2. Should the model be able to improve over time (self-learning)?

No, self-learning is not required. We prefer manual model updates when needed.

3. Should the interpreter support multiple languages in a single text input?

No, assume one language per input.

4. Do you want a confidence score along with the predicted language?

Yes, a confidence score (e.g., 92% sure it's Spanish) would be very helpful.

5. Should the interpreter be able to work offline?

No, online access is fine for now. Offline functionality is not a priority.

6. Is there a maximum input length the system should handle?

Yes, limit inputs to about 500 characters for text predictions.

7. Should the system support user-uploaded files (like .txt or .docx)?

Not for now — only direct text input (copy-paste or typing).

8. Should the system display a list of top 3 possible languages if it's unsure?

Yes, showing the top 3 predictions would be a good feature.

9. Will there be different user roles (admin, regular user)?

No, all users will have the same access — no role-based control needed.

10. Do you prefer a pre-trained machine learning model or a fully custom-built one?

We prefer using a pre-trained model to save time and resources.

GPS Map Camera
Hyderabad,Telangana,India
Gajularamaram, Hyderabad, 500055, Telangana, India
Lat 17.518936, Long 78.427358
12/06/2024 05:32 PM GMT+05:30
Note : Captured by GPS Map Camera

Client-3:

1. Will the system need to detect the script type (like Latin, Cyrillic, Devanagari)?

No, only language prediction is needed, not script identification.

2. Should the interpreter have a user authentication/login system?

No, no login required — the interpreter should be open for anyone to use.

3. What is the expected traffic/load — how many users per day?

We expect about 500–1000 users per day initially.

4. Will you need an API for third-party systems to use the language prediction?

Yes, an API endpoint would be helpful for integration later.

5. Should the system log user predictions for analysis?

Yes, log predictions anonymously (without saving user identity).

6. Are there any languages you specifically want to exclude from detection?

No, detect all major languages as long as they are in the supported list.

7. Should the model handle slang or informal text?

If possible, yes — basic slang handling would be good, but it's not critical.

8. What devices should be primarily targeted (desktop, mobile, tablet)?

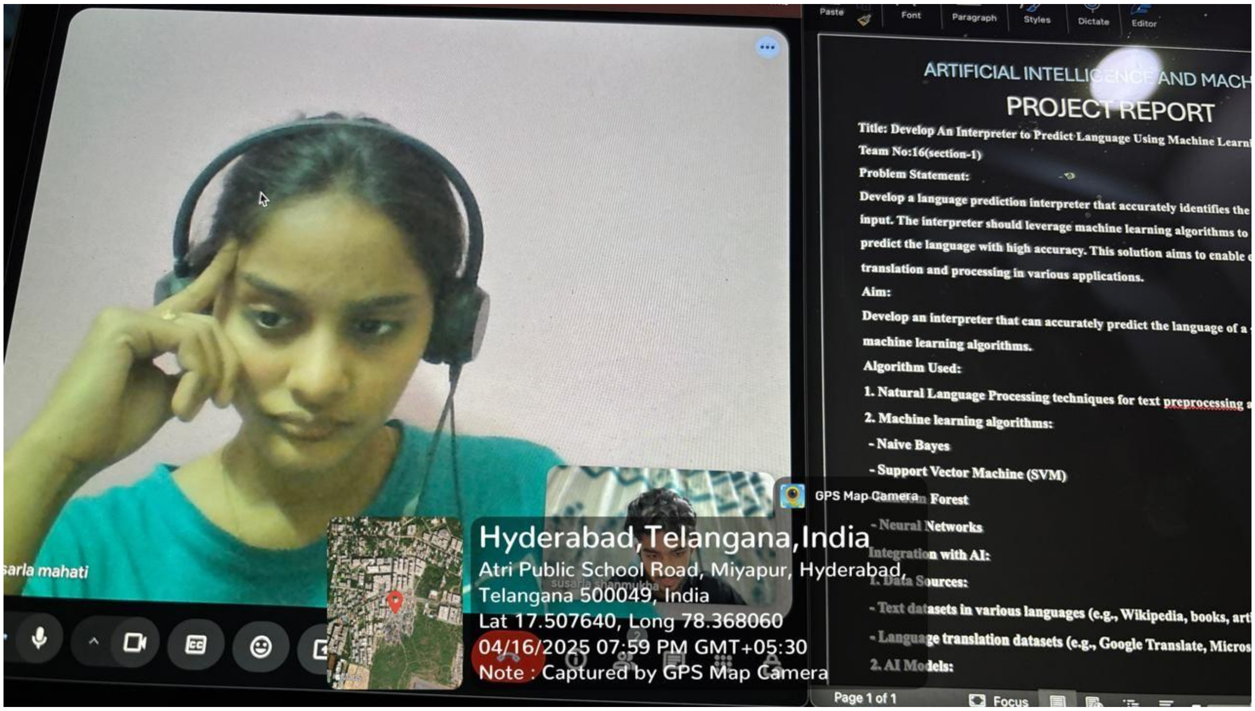
The interpreter should work well on both desktop and mobile.

9. Should we prioritize lightweight models for faster load time?

Yes, fast performance is very important — lightweight models are preferred.

10. Will there be a help or FAQ section for users?

Yes, a small FAQ page would be useful to guide users.



ARTIFICIAL INTELLIGENCE AND MACH

PROJECT REPORT

Title: Develop An Interpreter to Predict Language Using Machine Learni

Team No:16(section-1)

Problem Statement:

Develop a language prediction interpreter that accurately identifies the input. The interpreter should leverage machine learning algorithms to predict the language with high accuracy. This solution aims to enable translation and processing in various applications.

Aim:

Develop an interpreter that can accurately predict the language of a machine learning algorithms.

Algorithm Used:

1. Natural Language Processing techniques for text preprocessing

2. Machine learning algorithms:

- Naive Bayes
- Support Vector Machine (SVM)
- Forest
- Neural Networks

Integration with AI:

1. Data Sources:

- Text datasets in various languages (e.g., Wikipedia, books, art
- Language translation datasets (e.g., Google Translate, Micros

2. AI Models:

Hyderabad,Telangana,India
Atri Public School Road, Miyapur, Hyderabad,
Telangana 500049, India
Lat 17.507640, Long 78.368060
04/16/2025 07:59 PM GMT+05:30
Note : Captured by GPS Map Camera

Client-4:

1. Should the interpreter support right-to-left (RTL) languages like Arabic and Hebrew?

Yes, the system should properly handle RTL languages for correct display.

2. What should happen if the input is empty or too short to predict?

The system should show an error message like "Input too short for prediction."

3. Do you want the system to suggest corrections if the input is detected as a typo-heavy text?

No, no correction suggestions are needed - just predict based on the given input.

4. Should the interpreter provide examples of each supported language for users to view?

Yes, a small example sentence for each language would be a helpful addition.

5. Will there be a need for a dark mode or theme customization?

Not necessary at launch — light mode is fine for the first version.

6. How frequently do you plan to update the language model?

We plan to update it every 6 months or when a major improvement is available.

7. Should the system support mixed-language detection in the future?

Maybe later — not needed for the first release.

8. Should users be able to give feedback if the prediction is wrong?

Yes, a simple feedback option would be useful for improving the system.

9. Do you have a preferred cloud platform for hosting (AWS, Azure, GCP)?

We prefer to host it on AWS.

10. Should there be multilingual support for the app interface itself?

No, English-only interface is fine for now.

GPS Map Camera
Google
Hyderabad, Telangana, India
Dhivya Shree Shakthi Block-6, Dhivya Shree Shakthi, Mayuri Nagar, Miyapur, Hyderabad, Telangana 500049, India
Lat 17.514646°
Long 78.366854°
08/07/24 10:40 AM GMT +05:30