

## DAY – 20

**26 August 2025**

### What Is Used in My Project

#### 1 Project Type

My project is a **Web-Based Bilingual Chatbot System** developed to provide **irrigation-related information** to users through a simple chat interface.

It works as a **virtual assistant** for the **Punjab Irrigation Support and Management System (PISMS)**.

#### 2 Frontend (User Interface)

##### Technologies Used

- **HTML**
- **CSS**
- **JavaScript**

##### Purpose

The frontend is used to:

- Create the chatbot popup
- Take user input
- Display chatbot responses
- Allow language selection (English / Punjabi)

## **Files Used**

- index.html → Structure of chatbot
- style.css → Chatbot design and layout
- script.js → Sending user messages to backend and showing responses

## **3 Backend (Server Side)**

### **Technology Used**

- **Python**
- **Flask Framework**

### **Purpose**

The backend:

- Receives user queries
- Processes text using NLP
- Matches queries with dataset and PDF
- Sends the final answer back to frontend

### **Main File**

- app.py → Main server file

## **4 Dataset (Predefined Knowledge Base)**

### **File Used**

- dataset.txt

## Purpose

- Stores frequently asked questions and answers
- Used for quick and fixed responses
- Supports bilingual data

## Example

What is Jamabandi = Jamabandi is an official land record document.

Dataset is checked **first** for faster results.

## 5 PDF Processing (User Manual)

### File Used

- PRSC.pdf / PISMS User Manual

### Libraries Used

- **PyPDF2**

## Purpose

- Extract text from official user manual
- Answer procedural and detailed questions
- Reduce manual reading of PDF

## 6 Natural Language Processing (NLP)

### Libraries Used

- **NLTK**
- **Scikit-learn**

## **Purpose**

NLP is used to:

- Clean user input
- Remove stopwords
- Normalize text
- Improve matching accuracy

## **7 Machine Learning**

### **Models Used**

- **TF-IDF Vectorizer**
- **Logistic Regression**
- **Cosine Similarity**

## **Purpose**

- Convert text into numerical form
- Compare user query with dataset and PDF text
- Find the most relevant answer

## **8 Language Support (Bilingual Feature)**

### **Languages**

- English
- Punjabi

**Purpose**

- Improve accessibility for local users
- Provide responses in user's selected language

**9 Chatbot Logic****File Used**

- chatbot\_logic.py

**Purpose**

- Controls query flow
- Dataset matching
- PDF matching
- Confidence scoring
- Response selection

**Workflow Summary (Very Important)**

1. User types question
2. Language detected
3. Text cleaned using NLP
4. Dataset checked
5. If not found → PDF searched
6. Best answer selected
7. Response sent to user

## **Tools & Environment**

- Python Virtual Environment
- VS Code
- Flask Development Server
- Browser for testing

## **Advantages of Using These Technologies**

- Easy to maintain
- Low cost
- Scalable
- Beginner friendly
- Real-time response

## **Limitations**

- Domain-specific answers only
- Internet required
- No learning from new queries (currently)