***Project Title:***

Ice Cream Flavor Social Prediction

Members-

2420030423- J.Keerthana

2420030424- M.Sahithi

2420030520- Tejaditya

***Project Goal (Aim):***

To create an AI model that predicts which ice cream flavor will be the most popular based on **social media trends**, **seasons**, and **user preferences**.

* In **summer**, people might talk more about *mango* or *lemon* ice cream.
* In **winter**, *chocolate* or *coffee* flavors might be more popular.
* If a celebrity posts about *pistachio ice cream* on Instagram, suddenly it can trend.

We can collect this kind of **data** and train a model to **predict the next trending flavor**.

***Data We Can Use:***

We need two types of data:

1. **From Social Media** (Twitter, Instagram, Reddit, Google Trends)
   * Posts mentioning ice cream flavors.
   * Hashtags (#VanillaLove, #ChocolateIceCream).
   * Sentiment (positive or negative feelings about the flavor).
2. **From Datasets (public sources)**
   * Ice cream sales by season, location, and flavor.
   * Weather and temperature data.
   * Festivals or events (e.g., during Christmas, peppermint flavor might trend).

***Step-by-Step Process:***

**1. Data Collection**

* Use **Twitter API** to collect tweets about ice cream.
* Download datasets from **Kaggle** for ice cream sales.
* Use **Google Trends API** to see what flavors people are searching for.

Example data:

| **Date** | **Location** | **Flavor** | **Mentions** | **Sentiment** | **Temp (°C)** |
| --- | --- | --- | --- | --- | --- |
| 2025-06-12 | Delhi | Mango | 2300 | Positive | 35 |
| 2025-06-12 | Delhi | Chocolate | 1200 | Positive | 35 |

***2. Data Cleaning & Preprocessing:***

* Remove irrelevant text.
* Convert text to lowercase (e.g., "Chocolate" → "chocolate").
* Extract **keywords** for flavors.
* Use **Sentiment Analysis** (NLTK or TextBlob) to find whether people like or dislike a flavor.
* Encode non-numeric data (like “Positive” → 1, “Negative” → 0).

***3. Data Analysis (EDA):***

* Check which flavors are most talked about in which month.
* See if temperature affects flavor choice.
* Look at spikes in mentions after events or promotions.

***4. Model Building:***

We can use a **Classification Algorithm** because we’re predicting categories

Good algorithms to try:

* **Random Forest** (good for accuracy and easy to use)
* **Logistic Regression** (simple, good starting point)
* **Naive Bayes** (good for text-based predictions)

***Inputs to Model:***

* Month / Season
* Location
* Sentiment Score
* Temperature
* Number of social media mentions for each flavor

***Output from Model****:*

* The predicted trending flavor.

***5. Visualization***

* Use graphs to show:
  + Popularity trends over time.
  + Which flavor is most popular in which season.
  + Sentiment vs popularity.

***Tools & Technologies:***

* **Python**
* ***Libraries:***
  + Pandas, NumPy (for data handling)
  + Matplotlib, Seaborn (for graphs)
  + Scikit-learn (for machine learning)
  + NLTK/TextBlob (for sentiment analysis)
* **APIs**:
  + Twitter API
  + Google Trends API
* **Jupyter Notebook** for coding

***Example Prediction:***

**Input**:

* Month: July
* Location: Mumbai
* Top hashtags: #MangoIceCream, #SummerCool
* Sentiment Score: High positive
* Temperature: 33°C

***Output****:*

* Predicted Trending Flavor: **Mango**