**IFT 530**

**ADVANCED DATABASE MANAGEMENT SYSTEMS**

**FINAL PROJECT SUMMARY**

### **TOPIC: DATABASE MANAGEMENT FOR SENTIMENT ANALYSIS ON SOCIAL MEDIA**

**Group Number 16**

| **Name** | **ASURITE UserID** | **ASU ID** |
| --- | --- | --- |
| [Jahnavi Lasyapriya Vavilala](mailto:jvavilal@asu.edu) | jvavilal | 1231997652 |
| Sruthi Keerthana Nuttakki | snuttakk | 1232316477 |

**PROJECT SUMMARY**

**1. ABSTRACT AND USE OF THE TABLES**

Social media is an engaging two-way platform fostering communication amongst the customers and brands/ events. With social media gaining popularity in recent years, it can be leveraged to understand brands’ perception in the public. Sentiment analysis determines the opinion of customers allowing businesses to optimize their products/ services. Apart from businesses sentiment analysis is helpful in law enforcement, research, education, journalism, entertainment etc. Therefore, with sentiment analysis one can understand the trends, public opinion, engagement to reflect on decision making. Using Natural Language Processing (NLP), Machine Learning (ML) and text classification sentiments are grouped into positive, negative and neutral.

In the scope of this project, a database system will be created to handle volumes of data for efficient storage, organization and retrieval. A database system ensures high querying, allows cross platform analysis and facilitates easy updation.

In this project, there would be 8 tables and 2 fact tables as follows:

**Main Tables**:

1. **Users**: Information about social media users (anonymous identifiers).
2. **Posts**: Raw social media posts (content, metadata).
3. **Platforms**: Details about social media platforms.
4. **Hashtags**: Track hashtags used in posts.
5. **Keywords**: Identify key terms for sentiment analysis.
6. **Sentiment\_Scores**: Store sentiment scores for analyzed posts.
7. **Languages**: Store supported languages.
8. **Locations**: Record the geographical location of posts.

**Fact Tables**:

1. **Post\_Analytics**: Aggregate data for posts (e.g., likes, shares, comments).
2. **Sentiment\_Trends**: Store trends over time (e.g., average sentiment by day).

**2. TABLE SUMMARY**

1. **Users**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **User\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for users |
| **Username** | VARCHAR | 50 | NOT Null | - |  | Anonymized username |
| **Age** | INT | - | Optional | CHECK (Age >= 13) |  | Age of the user |
| **Gender** | CHAR | 1 | Optional | - |  | Gender of the user (M/F/O) |
| **Location\_ID** | INT | - | Optional | FK |  | Links to Locations table |

1. **Posts**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Post\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for posts |
| **User\_ID** | INT | - | NOT Null | FK |  | Links to Users table |
| **Platform\_ID** | INT | - | NOT Null | FK |  | Links to Platforms table |
| **Post\_Content** | TEXT | - | NOT Null | - |  | Stores social media post content |
| **Language\_ID** | INT | - | Optional | FK |  | Links to Languages table |
| **Post\_Date** | DATETIME | - | NOT Null | - |  | Date and time of the post |
| **Hashtags** | TEXT | - | Optional | - |  | Comma-separated list of hashtags |
| **Keywords** | TEXT | - | Optional | - |  | Comma-separated list of keywords |

**iii. Platforms**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Platform\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for platforms |
| **Platform\_Name** | VARCHAR | 50 | NOT Null | - |  | Name of the social media platform |

**iv.Hashtags**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Hashtag\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for hashtags |
| **Hashtag\_Text** | VARCHAR | 100 | NOT Null | UNIQUE |  | Hashtag text |

**v.Keywords**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Keyword\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for keywords |
| **Keyword\_Text** | VARCHAR | 100 | NOT Null | UNIQUE |  | Keyword text |

**vi.Sentiment\_Scores**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Score\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for sentiment scores |
| **Post\_ID** | INT | - | NOT Null | FK |  | Links to Posts table |
| **Sentiment\_Score** | DECIMAL | 3,2 | NOT Null | CHECK (Sentiment\_Score BETWEEN -1.00 AND 1.00) |  | Sentiment score (-1 to 1) |

**vii.Languages**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Language\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for languages |
| **Language\_Name** | VARCHAR | 50 | NOT Null | UNIQUE |  | Language name (e.g., English, Spanish) |

**viii.Locations**

| **Column Name** | **Data Type** | **Size** | **Null / NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Location\_ID** | INT | - | NOT Null | UNIQUE | PK | Unique identifier for locations |
| **Location\_Name** | VARCHAR | 100 | NOT Null | - |  | Name of the location |

**ix.Post\_Analytics (Fact table)**

| **Column Name** | **Data Type** | **Size** | **Null/NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Analytics\_ID** | INT | - | Not Null | UNIQUE | PK | Unique identifier for analytics record |
| **Post\_ID** | INT | - | Not Null | FK |  | Links to Post table |
| **Likes\_count** | INT | - | Not Null | DEFAULT 0 |  | Number of likes on the post |
| **Shares\_Count** | INT | - | Not Null | DEFAULT 0 |  | Number of shares on the post |
| **Comments\_Count** | INT | - | Not Null | DEFAULT 0 |  | Number of comments of the post |
| **Views\_Count** | INT | - | Not Null | DEFAULT 0 |  | Number of Views on the post |
| **Engagement\_Rate** | DECIMAL | 5,2 | Not Null | DEFAULT 0 |  | Calculated engagement rate |
| **Analysis\_Date** | DATETIME | - | Not Null | **-** |  | Date and time of analysis |

**x.Sentiment\_Trends (Fact Table)**

| **Column Name** | **Data Type** | **Size** | **Null/NOT Null** | **Constraint** | **PK** | **Purpose of the Column** |
| --- | --- | --- | --- | --- | --- | --- |
| **Trend\_ID** | INT | - | Not Null | UNIQUE | PK | Unique identifier for trend record |
| **Platform\_ID** | INT | - | Not Null | FK |  | Links to Platforms table |
| **Date** | DATE | - | Not Null | - |  | Date of the trend analysis |
| **Avg\_Sentiment** | DECIMAL | 3,2 | Not Null | CHECK(Avg\_Sentiment BETWEEN -1.00 AND 1.00) |  | Average sentiment score |
| **Post\_Count** | INT | - | Not Null | DEFAULT 0 |  | Total number of posts analyzed |
| **Positive\_Ratio** | DECIMAL | 5,2 | Not Null | CHECK (Positive\_Ratio BETWEEN 0 AND 100 |  | Percentage of positive posts |
| **Negative\_Ratio** | DECIMAL | 5,2 | Not Null | CHECK (Negative\_Ratio BETWEEN 0 AND 100) |  | Percentage of negative posts |
| **Neutral\_Ratio** | DECIMAL | 5,2 | Not Null | CHECK (Neutral\_Ratio BETWEEN 0 AND 100) |  | Percentage of neutral posts |