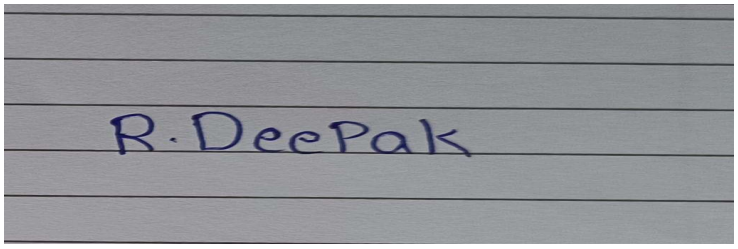
Annexure3b- Complete filing

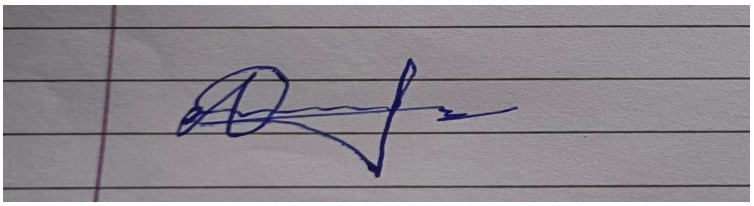
INVENTION DISCLOSURE FORM

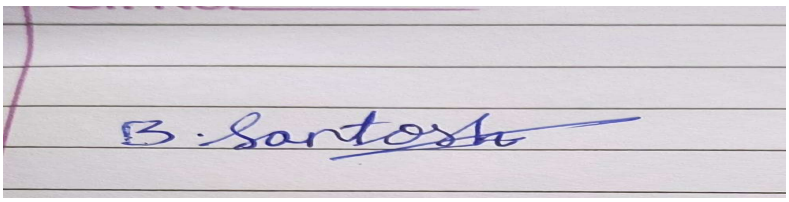
Details of Invention for better understanding:

1. TITLE: Sentimental Analysis of Social Media Presence AI-Based Emotion Detection & Reputation Insights System

2. INTERNAL INVENTOR(S)/ STUDENT(S): All fields in this column are mandatory to be filled

A. Full name	Raavi Deepak
Mobile Number	7893843943
Email (personal)	Raavideepak5@gmail.com
UID/Registration number	12311018
Address of Internal Inventors	Lovely Professional University, Punjab-144411, India
Signature (Mandatory)	

B. Full name	Shaik Amaan
Mobile Number	7207404271
Email (personal)	Shaikamaan885@gmail.com
UID/Registration number	12313285
Address of Internal Inventors	Lovely Professional University, Punjab-144411, India
Signature (Mandatory)	

C. Full name	Santosh Sai Raj
Mobile Number	9492802727
Email (personal)	santoshsairaj77@gmail.com
UID/Registration number	12310454
Address of Internal Inventors	Lovely Professional University, Punjab-144411, India
Signature (Mandatory)	

EXTERNAL INVENTOR(S): (INVENTORS NOT WORKING IN LPU)

A. Full name	
Mobile Number	
Email	

Address of External Affiliations	
Signature (Mandatory)	

***For External Inventors*, NOC (No Objection Certificate) from the affiliated institute/university/Industry/lab etc. is mandatory for each individual inventor and their respective topic. For NOC, format is attached below.**

(FOR ADDITIONAL INVENTORS, PLEASE ADD ROWS)

3. DESCRIPTION OF THE INVENTION:

A. PROBLEM ADDRESSED BY THE INVENTION:

- **Manual Sentiment Checking is Impossible**
  Manual Sentiment Checking is Impossible. A million users are getting thousands of comments. It is laborious and inefficient to manually read the comments
- **Hidden Negative Sentiments**
  No one hears the toxic comments, subtle hate, sarcasm, or bullying among the noise.
- **Brand / Personal Image Unclear**
  Users don't know their overall digital reputation score
- **No Real-Time Emotion Tracking**
  Current platforms lack the feature of instant emotional analysis of new comments or posts.
- **Fragmented Information Sources**
  Comments are scattered across Instagram, YouTube, X (Twitter), Facebook, etc. There is no single system that merges them.

The Sentimental Analysis System solves all these problems by constantly fetching and analysing social media data from multiple platforms, then converting it into clear sentiment insights, emotion categorisations, and a useful reputation score.

B. OBJECTIVE OF THE INVENTION (Provide minimum two)

1. **To Automatically Understand Public Emotion From Social Media Content.**
   The main point of the invention is to develop an AI system that can spontaneously read and comprehend people's emotions when they comment, post, or react on platforms such as Instagram, YouTube, X, and Facebook. The system is the one that, in a flash, decides whether the general feeling is positive, negative, or neutral.

.

2. **To Detect Deeper Emotions Such as Support, Hate, Anger, Motivation, or Sarcasm**. This system doesn't just look at whether a post is positive or negative. It tries to understand the real feeling behind the words. It can catch support, anger, hate, motivation, sarcasm, sadness, trolling, or bullying—things that normal sentiment tools often miss. The goal is to read the text the way a real person would and understand the true emotion in it**.**

3. **To Generate an Easy-to-Understand Reputation Score**
   The system comes up with a straightforward score that depicts a person's overall online image based on the emotional reactions and engagement.

4. **To Combine Social Media Data Into One Clear Platform**
   By gathering the reactions from different platforms, the invention presents all the sentiment information in one unified dashboard.

5. **To Help Users Make Better Decisions Based on Sentiment Insights**
   By knowing which content brings positivity or negativity, the system will be instrumental for users to improve their online presence and interaction strategies

C. STATE OF THE ART/ RESEARCH GAP/NOVELTY: Describe your invention fulfil the research gap?

Sr. No.	Patent I'd	Abstract	Research Gap	Novelty
1.	US20160232043A1	Describes a sentiment analysis system using machine learning to Classify t h e text into positive, negative, and neutral categories. It processes user text and applies basic ML models for prediction	Focuses only on **simple sentiment classification**. No emotion layers, no toxicity detection, no platform-wise aggregation, and no reputation scoring.	Your invention integrates **sentiment** + **emotion** + **sarcasm** + **toxicity** + **reputation score**, offering a complete multi-dimensional social media analysis system
2.	US20190291721A1	A deep neural network model that identifies emotional categories such	Limited to **emotion detection only**. No sentiment trends, no cross-	Your system combines **emotions** + **sentiment trends** + **user**

		as anger, joy, fear, and sadness from text inputs. Works mainly on emotion detection..	platform data fusion, no commenter behaviour analysis	**influence + engagement scoring** into a unified social media presence analysis tool.
3.	US20200345612A1	Patent provides a method for detecting toxic, abusive, or harmful text using neural networks and classification layers..	Focuses only on **toxicity**. Does not integrate sentiment classification, emotional tone, sarcasm detection, or reputation tracking	Your invention merges **toxicity + sentiment + emotion + sarcasm** and generates a **dynamic reputation score**, creating a more comprehensive analysis engine.

D. DETAILED DESCRIPTION:

1. Overview

The Sentimental Analysis System is an AI/ML end-to-end platform that on its own is able to gather social media text and reactions, cleanse and process the data into features that machines can understand, run the sentiment and emotion models and by the user interface (web/mobile) provide clear, actionable insights (sentiment breakdowns, emotion labels, reputation score, and alerts) to the user).

. It has been architectured to be account, page, and platform scalable as well as easily translatable to different languages and networks

2. System Components & Design

I. System Requirements

1. Hardware Requirements

Minimum (for development / running models locally)

- **Processor:** Dual-Core CPU (Intel i3 / AMD equivalent)
- **RAM:** 6 GB
- **Storage:** 10–20 GB free (for datasets, Python, libraries, reports)
- **Graphics:** Integrated graphics (no special GPU needed)
- **Others:**
    - Stable internet (for weather API / package installation)
    - Keyboard, mouse, basic display

Recommended (for smoother ML work)

- **Processor:** Intel i5 / Ryzen 5 or higher
- **RAM:** 8 GB or more
- **Storage:** 50+ GB free (for multiple datasets, models, backups)
- **Optional:**
  - NVIDIA GPU (if later add deep learning, not required now)

2. Software Requirements

Operating System

- Windows 10 / 11 **(most likely)**
- or Linux (Ubuntu)
- or macOS

Development Tools

- **Python** 3.8 or above
- **IDE / Editors:**
  - Jupyter Notebook / JupyterLab
  - VS Code / PyCharm (optional but helpful)

3. Python Libraries Required

Core Libraries (for ML + data handling)

- pandas – data loading and preprocessing
- numpy – numerical operations
- scikit-learn – ML models (Logistic Regression, Naive Bayes, Random Forest, train-test split, metrics)
- matplotlib – basic plots (yield, rainfall, risk charts)
- seaborn – attractive visualizations (heatmaps, distributions)
- nltk, spacy — tokenization, lemmatization, stopwords
- joblib / pickle — model persistence
- transformers (Hugging Face) — contextual embeddings / BERT models

4. Frontend Technologies (for Web UI / Dashboard)

Basic Frontend Stack

- **HTML5** – structure of web pages
- **CSS3** – styling (colors, layout, responsive design)
- **JavaScript** – basic interactivity and input validation
- **React.js -** for SPA

Frameworks / Libraries (optional but good for projects)

- **Bootstrap** – for quick responsive UI, buttons, forms, cards

- **Chart.js** or **Plotly.js** – to show graphs in browser (yield, rainfall, risk level)
- **React.js** (optional advanced) – if you want a modern single-page app

Backend Connection

- Frontend (HTML/CSS/JS) communicates with:
  - **Flask / FastAPI** backend in Python
  - Backend runs ML model and sends predictions (best crop, risk level, precautions) in JSON format.

II. Design

1. **Data Sources**

- **Platform APIs:** Instagram, YouTube, X (Twitter), Facebook, LinkedIn.
- **Scraped Content:** Public comments, captions, replies, and post text.
- **Engagement Signals:** Likes, reactions, retweets/shares, view counts, reply counts.
- **Metadata:** Timestamps, language, user handle, platform, post id, follower counts.
- **optional:** Uploaded CSV of historical comments or third-party sentiment datasets for training.

2. **Data Processing & Feature Engineering**

- **Cleaning:** remove URLs, normalize emojis, strip HTML, remove boilerplate signatures.
- **Normalization:** lowercasing, expand contractions, mapping slang/short forms, handling repeated letters.
- **Emoji & Reaction Mapping:** convert emojis and reaction types into sentiment features
- **Language Handling:** detect language; route to language-specific pipelines or translation
- **Derived Features:** sentiment-bearing n-grams, punctuation intensity (!!!), capitalized word ratio, sarcasm indicators (e.g., interjections + punctuation), time-of-day posting pattern, engagement ratio (replies/views)

3. **Machine Learning Modules**

The system uses different machine learning models to understand the meaning, emotion, and overall sentiment behind social media comments. Each model plays its own role in improving accuracy and uncovering the deeper emotional patterns hidden in the text.

a) **Sentiment Classification (Positive / Negative / Neutral):**

- Models: Logistic Regression/ Naive Byes / Random Forest.

- Input: Cleaned social media comments/ Extracted features (TF-IDF / embeddings).
- Output: Sentiment label: *Positive / Negative / Neutral.*
  Confidence score for each prediction

b) **Emotion Detection Model (Deeper Emotions):**

- Support
- Anger
- Hate / Toxic
- Motivation
- Sarcasm
- Appreciation
- Sadness.

The model learns from authentic social media language, emojis, slang, and short forms to understand hidden emotions.

Purpose:

- Bullying detection on the net.
- Recognizing sarcastic or masked negative commentary.
- Spotting genuine support or appreciation.

c) **Sarcasm & Toxicity Classification:**

Sarcasm often looks positive on the surface, but the real meaning is negative. Toxic comments, on the other hand, include things like hate, abuse, harassment, or generally rude language.

One model alone is responsible for the following tasks:

- Setting the target audience for the most harmful or insulting comments
- Finding trolling or spam identity
- Enhancing the accuracy of sentiment predictions

This model is important because many comments online are not straightforward.

4. **Reputation Score & Trend Detection Model**

- The model Combines
  - Sentiment results
  - Emotional intensity
  - Engagement quality
  - Frequent negative spikes

- Positive support trend

The output is a **Reputation Score** that represents how the public views the person or brand

It also analyses how sentiment changes over days/weeks to identify:

- Growth in support
- Sudden negativity
- Engagement improvement
- Audience mood changes.

5. **User Interface & Integration**
   - **Mobile / Web App** for General Users:
     - Quick upload or text-input option for posts, comments, or statements.
     - Dashboard showing recommended Sentiment result, Emotional Score, and Reputation Score, and Daily/weekly sentiment trend.
   - **Alerts:**
     - Negative spikes/Hate speech/Sudden reputation drop.
   - **Data Export & Reporting**:

. Users can fix the model's mistakes by marking which results are wrong. These corrections are saved and later used to retrain the system so it becomes more accurate over time.

3. Working

1. **Data Collection & Update**
   - System polls APIs or receives webhooks periodically; users can upload historical datasets
2. **Model Inference**
   - For each comment: cleaning → embedding → model predictions (sentiment, emotion, toxicity).
3. **Aggregation & Scoring**
   - Aggregate comment-level results at post/account level; compute reputation score and trend metrics.
4. **Recommendation Generation**
   - Produce weekly action plan: top negative threads, posts to boost, moderation priorities.
5. **Feedback & Learning**

Users can correct model outputs (mark false positives/negatives); corrections are stored and used for periodic retraining**.**

E. RESULTS AND ADVANTAGES:

RESULTS

- Accurate classification of large volumes of comments into sentiment and emotions.
- Timely detection of negative sClear reputation metric and platform-wise sentiment breakdown.

- Actionable suggestions for content improvement and moderation.

ADVANTAGES

1. **Data-Driven Reputation Management**
    - replaces guesswork with measurable insights..
2. **Early Warnings for Crisis**
    - detects sudden negativity to allow quick responses.
3. **Efficient Moderation**
    - prioritizes comments needing human review.
4. **Location-Specific and Personalized**
    - unified picture of online presence.
5. **Scalable & Updatable**

Feature / Parameter	Logistic Regression	Naive Bayes	Random Forest	Transformer(BERT)
Model Type	Linear classifier	Probabilistic classifier	Ensemble (multiple decision trees)	Deep contextual model
Works Best For	Simple patterns	Small datasets & noisy data	Complex Patterns	Context, sarcasm
Data Used	Cleaned comments + TF-IDF feature	Same features as Logistic Regression	Same features with higher variance tolerance	Word embeddings + contextual tokens
Training Speed	Fast	Very fast	Moderate	Slow
Prediction Speed	Fast	Fast	Moderate	Highest
Accuracy (Typical)	Moderate	Moderate	High	Highest
Interpretability	High	High	Medium	Low
Handles slang context	Poor	Limited	Good	Excellent
Overfitting Risk	Low	Moderate	High (if not tuned)	High
Suitable For Sentiment Prediction	Yes (baseline)	Yes (probabilistic)	Yes (best performance)	Yes(best for deep Context)

Suitable For Emotion Detection	Limited	Limited	Good for multi-class	Excellent (best)
Outputs Provided	Class label + probability	Class label + probability	Class label + feature importance+ probability	feature importance Class label + confidence + contextual embeddings
Strengths	Easy to train, interpretable	Fast, works well on limited data	High accuracy, handles complex patterns	Best understanding, handles sarcasm,

- models retrain with new data and slang.

MODEL COMPARISION TABLE

F. EXPANSION:

To prevent competitors from making minor modifications, the patent can cover:

1. **Data Fusion Framework**
   - This explains how we take information from different platforms—like text, emojis, and reactions—and combine them into one clean, organized data format. Instead of treating each platform separately, everything is merged into a single, easy-to-analyze feature set so the system understands the full meaning behind a user's comment.

2. **Hybrid ML Architecture**

   This describes how we mix traditional machine-learning models with modern transformer models. By arranging them in the right order, we get better accuracy, stronger context understanding, and overall smarter predictions. It's like combining old-school reliability with new-age intelligence..

3. **Risk Scoring & Advisory Generation Logic**
   - This is the logic that converts continuous ML outputs into a final rating. It takes the system's raw predictions and converts them into a simple, easy-to-understand reputation score. Based on that score, the system also provides clear suggestions or warnings that users can act on..

4. **Feedback-Loop Learning Mechanism**

   This explains how the system improves over time. When users correct the they're feed back into the training pipeline so the model becomes more accurate with every upd

- This covers how the system identifies sudden increases in activity or risk in real time and alerts the user immediately. It also describes how it gives step-by-step guidance—for example, suggesting the right crop, the best sowing time, how much to irrigate, or how to handle pests—using local language and timely notifications.

G. WORKING PROTOTYPE/ FORMULATION/ DESIGN/COMPOSITION:

A working prototype for the Sentimental Analysis System has been made using Python and machine learning techniques. The prototype can analyse the comments from social media in real time and produce the following insights: sentiment, emotion, toxicity, and reputation

- **Data Layer:**

The present prototype relies on CSV-based datasets containing:

- Social media comments and replies
- Captions of posts and hashtags
- Emojis and expressive symbols
- The number of likes, replies, shares, and views • User metadata (followers, verification, account age)
- Sentiment/emotion labels for training (positive, negative, neutral, anger, hate, sarcasm, joy, etc.)

- **ML Layer:**

- **The prototype embodies three main models**

  - Preprocessing and feature scaling (cleaning text, removing noise, tokenization, vectorization)
  - Classification models for sentiment labels (positive, negative, neutral).
  - Emotion recognition models (anger, sarcasm, support, joy, hate, etc.).
  - Clustering model to group similar comment types, trends, and audience behavior.

- **Application Layer:**

  - Jupyter Notebook / script interface where a user uploads a dataset (e.g., cropdata.csv) and receives:

    - **Sentiment classification** (positive / negative / neutral),
    - **Emotion category** (anger, sarcasm, support, joy, hate, etc.),
    - **Toxicity level** (non-toxic / toxic / abusive).
    - Reputation score
    - Suggestions for improving online presence

Next Steps in Prototype Development:

- Real-time API integration for Instagram, YouTube, X, Facebook..
- Creation of a user-friendly web/mobile UI.
- Implementation of multilingual sentiment analysis (English, Hindi, Telugu, Tamil, etc.)
- Add **live negative spike alerts**
- Extension to various social media platforms.

Machine Learning Models Implemented:

1. Logistic Regression Model (LogisticReg.ipynb)

Purpose:
To use text features for classifying comments as positive, negative, or neutra.

Dataset Used:

- Comment text
- Emojis
- Hashtags
- Engagement metrics
- Basic user metadata
- Sentiment labels (positive, negative, neutral)

Pipeline Steps:

- Load CSV file using pandas.
- Clean text and remove noise
- Encode labels.
- Convert text into TF-IDF vectors.
- Split into train–test sets.
- Train Logistic Regression model.
- Evaluate using:
    - Accuracy
    - Confusion Matrix
    - Classification Report

Role in Crop Guidance System:

- Baseline sentiment Classifer.
- Fast and interpretable.
- Good for comparison with advanced models.

2. Naive Bayes Model (NaiveBaiyes.ipynb)

Purpose:
To quickly give a probabilistic estimation of the sentiment or toxicity in the text.

Dataset Used:
The same as for Logistic Regression, with the addition of emojis and short-form text..

Pipeline Steps:

- Load and clean dataset.
- Clean Comments.
- TF-IDF or Count Vectorizer.
- Train Gaussian or Multinomial NB.
- Encode labels.
- Evaluate using:
    - Accuracy
    - Confusion Matrix
    - Probability outputs
    - Classification Report

Role in the Sentiment System:

- Extremely good for short and noisy social media comments.
- Provides probability scores for confidence estimation.
- Appropriate for large volume of comment processing..

3. Random Forest Model (RandomForest.ipynb)

Purpose:
To employ ensemble learning for pinpointing sentiment/emotion more accurately.

Dataset Used:
Complete set of cleaned comment text.

Extracted features(TF-ID, embeddings, emoji indicators)

Sentiment/emotion labels

Pipeline Steps:

- Import RandomForestClassifier / RandomForestRegressor.
- Train model using bootstrapped decision trees.
- Predict outputs for best crop or yield category.
- Evaluate using:
    - Accuracy
    - Feature Importance

- Confusion Matrix
- Classification Report

Role in the Sentiment System:

- Stable and robust enough for real-world social media data..
- Can handle slang, mixed emotions, irregular language..
- Demonstrates which features (words/emojis) are most significant G. EXISTING DATA

G. EXISTING DATA:

- Publicly available sentiment datasets (Sentiment140, SemEval, GoEmotions.
- Toxicity datasets (Jigsaw Toxic Comment Dataset).
- Sarcasm datasets (SARC and curated irony datasets).
- Exported Instagram/YouTube/X comments (CSV format)
- Labeled samples created manually for project testing

These datasets are utilized to **train and validate and improve model performance**

The system is designed to include new data continuously, improving accuracy over time

4. USE AND DISCLOSURE (IMPORTANT): Please answer the following questions:

A. Have you described or shown your invention/ design to anyone or in any conference?	YES ( )	NO (X)
B. Have you made any attempts to commercialize your invention (for example, have you approached any companies about purchasing or manufacturing your invention)?	YES ( )	NO (X)
C. Has your invention been described in any printed publication, or any other form of media, such as the Internet?	YES ( )	NO (X)
D. Do you have any collaboration with any other institute or organization on the same? Provide name and other details.	YES ( )	NO (X)
E. Name of Regulatory body or any other approvals if required.	YES ( )	NO (X)

5. Provide links and dates for such actions if the information has been made public (Google, research papers, YouTube videos, etc.) before sharing with us. **NA**

6. Provide the terms and conditions of the MOU also if the work is done in collaboration within or outside university (Any Industry, other Universities, or any other entity). **NA**

7. **Potential Chances of Commercialization.**

Yes, the Sentiment Analysis has strong potential for commercialization because:

- Online reputation management is rapidly growing.
- Brands, influencers, and creators require audience sentiment tools.
- Companies can utilizet for:
- Customer feedback Monitoring
- Brand reputation tracking
- Toxicity filtering
- Marketing Sentiment analysis
- Possible Business models:
    - **B2B APIs** ,
    - Saas subscription plans,
    - Social media monitoring dashboards.
    - Enterprise emotion intelligence software

8. **List of companies which can be contacted for commercialization along with the website link.**

A. Hootsuite

Website: https://www.hootsuite.com/

B. Sprout Social

Website: https://sproutsocial.com/

C. UPL Ltd. (United Phosphorus Limited)

Website: https://www.upl-ltd.com/

D. Brand watch

Website: https://www.simplify360.com/

E. Melt Water

Website: https://www.meltwater.com/en

F. Agorapulse

Website: https://www.agorapulse.com/

9. Any basic patent which has been used and we need to pay royalty to them.

10. FILING OPTIONS:

Provisional Patent Filing:

I am considering filing a provisional patent to establish an early priority date for my invention. Since I have a well-defined algorithm and a basic working prototype, this filing will give me 12 months to further refine and develop the technology before converting it into a complete patent. This approach is ideal for me because my AI model, skin diagnostics method, and allergy detection process are unique but still evolving.

11. **KEYWORDS:**

- Sentiment Analysis System
- Social Media Emotion Analysis
- Toxicity Detection
- Reputation Scoring Engine
- Machine Learning in Social Media
- Sarcasm Detection
- Multi-Platform Comment Analysis
- NLP-Driven Reputation Monitoring

(Letter Head of the external organization)

NO OBJECTION CERTIFICATE

This is to certify that University/Organization Name or its associates shall have no objection if Lovely Professional University files an IPR (Patent/Copyright/Design/any other…….) entitled "……." including the name(s) of, ...…as inventors who is(are) student(s)/employee(s) studying/ working in our University/ organization.

Further Name of the University/Organization shall not provide any financial assistance in respect of said IPR nor shall raise any objection later with respect to filing or commercialization of the said IPR or otherwise claim any right to the patent/invention at any stage.

(Authorised Signatory)