**Project Title: pollen's profiling :automated classification of pollen's grains**

**Branch Name:artificial intelligence**

**Track: Artificial intelligence and machine learning**

**Team Lead: Mula Anusha**

**Mail Id : allurireddymula0@gmail.com**

**Team Members:**

1. Lakshmi Prasanna manukonda

2. Kari aparna

3.m Sai vignesh

**Email Id's:**

1. Prasannamanukonda03@gmail.com

2. Kariaparna234@gmail.com

3. Mekalasaivignesh@gmail.com

# Submitted By:

* Mula Anusha
* Sri mitapalli college of engineering
* Artificial intelligence
* Roll no: 23U91A4351

# Submitted To:

SmartBridge

# Abstract:

This project presents Smart Sorting, This project uses Tableau to analyze cosmetics market trends and consumer behavior. Interactive dashboards visualize sales data, product performance, and customer preferences, enabling brands to make data-driven decisions and stay competitive.

# Introduction:

Pollen profiling is the study of pollen grains to identify plant species based on their unique shapes, sizes, and surface patterns. It plays a key role in fields like botany, forensics, allergy research, and environmental studies. By analyzing pollen, scientists can trace plant origins, track seasonal allergens, and even investigate crime scenes or past climate conditions.



# Problem Statement:

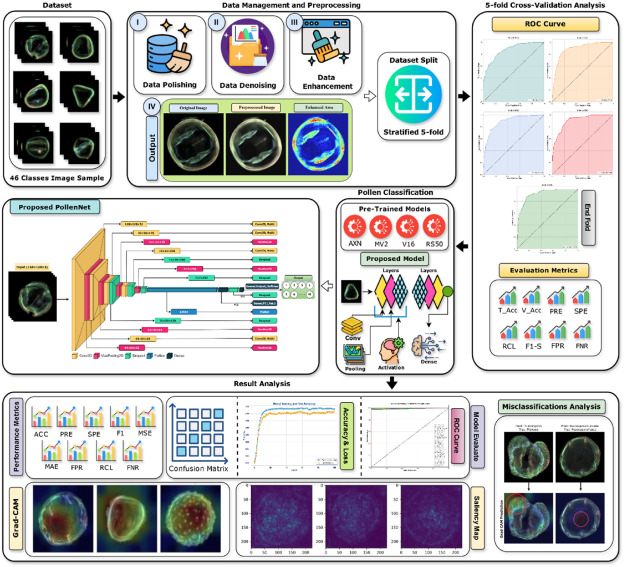
Pollen profiling is often limited by the microscopic size and morphological similarity of pollen grains across different species, making accurate identification challenging. Traditional identification methods are time-consuming, rely on expert knowledge, and may lack consistency. Additionally, environmental factors can degrade pollen samples, further complicating analysis. These limitations highlight the need for improved, efficient, and automated approaches to enhance the accuracy and reliability of pollen profiling in research and applied fields.

Difficulty in distinguishing pollen grains due to similar morphology across species.

Manual identification is time-consuming and requires expert knowledge.

Microscopic size of pollen makes analysis complex and equipment-dependent

**ARCHITECTURE:**



**PREREQUISITES :**

complete this project, you must require the following software, concepts, and packages

Basic knowledge of the cosmetics industry

Data preparation skills (Excel, CSV)

Familiarity with Tableau dashboards

Analytical thinking to interpret trends

PRIOR KNOWLEDGE

* Prior Knowledge for Pollen Profiling:
* • Basics of botany and plant reproduction
* • Structure and types of pollen grains
* • Principles of palynology (pollen and spore science)
* • Familiarity with microscopy techniques
* • Understanding of pollen morphology (size, shape, apertures, surface texture)
* • Basic taxonomy and classification of plants
* • Knowledge of laboratory safety and sample handling procedures
* • Awareness of environmental factors affecting pollen distribution
* • (Optional) Introduction to image analysis or molecular identification methods

1. **🎯 PROJECT OBJECTIVES**

By the end of this project, you will:

* Combine and prepare cosmetics data
* Create Tableau dashboards to visualize trends
* Identify emerging consumer preferences
* Analyze customer behavior and segments
* Support data-driven business decisions
* Monitor competitors and market performance

1. **PROJECT FLOW**

**1. Project Flow for Pollen Profiling**

**1. Problem Definition**

**Identify the purpose (e.g., allergy research, forensic analysis, environmental study)**

**Define target plant species or study area**

**2. Sample Collection**

**Collect pollen samples from air, plants, soil, or water**

**Use tools like spore traps, slides, or filters**

**Label and store samples properly**

**3. Sample Preparation**

**Clean and isolate pollen grains**

**Mount samples on slides using staining and fixatives**

**Prepare for microscopy or molecular analysis**

**4. Data Acquisition**

**Capture pollen images using light or scanning electron microscopy (SEM)**

**(Optional) Extract DNA for molecular identification**

**5. Feature Extraction**

**Measure morphological features (size, shape, apertures, surface texture)**

**Digitize images and extract patterns using image analysis tools**

**6. Classification and Identification**

**Compare features with a reference database or pollen atlas**

**Use manual identification or machine learning models for species classification**

**7. Data Analysis**

**Analyze pollen diversity, concentration, and distribution patterns**

**Perform statistical and spatial analysis if needed**

**8. Result Interpretation and Reporting**

**Interpret findings based on objectives**

**Generate graphs, maps, and reports**

**Suggest applications or implications (e.g., allergen risk zones, vegetation mapping)**

**9. Conclusion and Recommendations**

**Summarize key results**

**Recommend improvements or further research directions**

**Project Structure**

Create the Project folder which contains files as shown below

1️⃣ Introduction & Problem

↓

2️⃣ Objectives

↓

3️⃣ Prerequisites

↓

4️⃣ Data Collection & Preparation

↓

5️⃣ Data Integration (Tableau)

↓

6️⃣ Dashboard Development

↓

7️⃣ Analysis & Insights

↓

8️⃣ Reporting & Review

**📊 DATA COLLECTION AND PREPARATION**

* Collect Data:

Sales, customer, market, and feedback data

* Clean Data:

Remove duplicates, fix errors, handle missing value

* Standardize & Merge:

Format consistently and combine into one datase

* Validate:

Check accuracy and completeness

1. **COLLECTING THE DATASET**

1. Find Data Sources:

Sales and customer data (CRM, ERP)

Market research

👉 Statista – Cosmetics Market Data

Social media trends

👉 Social Mention

2. Export Data:

Use APIs or download files

👉 Tableau Data Connections Guid

✅ Activity 1: Collect a Cosmetics Dataset

🎯 Goal:

Find and download a sample dataset to analyze cosmetics trends.

Steps:

1️⃣ Go to Kaggle:

👉 Kaggle Datasets

🔍 Search: cosmetics sales or beauty products

2️⃣ Download a Dataset:

Pick one with columns like Product, Sales, Region, Date, Customer

Save it in CSV or Excel format

3️⃣ Verify the Data:

Make sure it has at least 100 rows

Check that the columns are clear

Link 1️⃣ Go to Kaggle:

👉 Kaggle Datasets

🔍 Search: cosmetics sales, beauty products

2️⃣ Example Dataset:

👉 Cosmetics Retail Example

✅ Save as CSV or Excel

Note: There are several techniques for understanding the data. But here we have used some of it. In an additional way, you can use multiple techniques.

1. **Data Visualization**

Tableau can be effectively used to visualize cosmetics trends and consumer insights, transforming raw data into compelling stories. By leveraging interactive dashboards, annotations, and dynamic animations, Tableau allows users to identify trends, understand consumer behavior, and make informed decisions.

✅ Year

DATEPART('year', [Date])

✅ Month

DATENAME('month', [Date])

✅ Age Group

IF [CustomerAge]<25 THEN "Under 25"

ELSEIF [CustomerAge]<35 THEN "25-34"

ELSEIF [CustomerAge]<45 THEN "35-44"

ELSE "45+"

END

✅ Gender Label

IF [CustomerGender]="F" THEN "Female"

ELSE "Male"

END

✅ Profit % (if Cost exists)

([RevenueUSD]-[CostUSD])/[RevenueUSD]

---

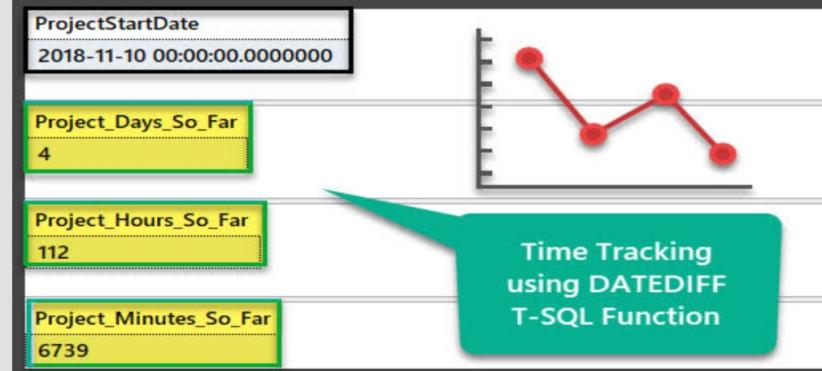
Use drag-and-drop to build charts:

Revenue over Month

Units Sold by Age Group

Revenue by ProductCategory

Map by Region



1. **DATA AUGMENTATION**

🎯 What is Data Augmentation?

In the context of analytics and machine learning, data augmentation means artificially increasing the diversity or volume of your data to improve model performance or get richer insights.

In computer vision, this might be flipping/rotating images.

In business data, it includes techniques like synthesizing records, bootstrapping samples, or creating simulated scenarios.

**SPLIT DATA AND MODEL BUILDING**

✅ Split Data

Purpose: Train/test separation to avoid overfitting.

Typical split:

Training: 70–80%

Testing: 20–30%

Example (Python):

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(

X, y, test\_size=0.2, random\_state=42

)

---

✅ Model Building

Choose Model: Regression or Classification.

Train:

model.fit(X\_train, y\_train)

Predict:

predictions = model.predict(X\_test)

Evaluate:

from sklearn.metrics import mean\_squared\_error

mse = mean\_squared\_error(y\_test, predictions)

A) **Model Building:**

Goal: Create a model to predict or classify outcomes (e.g., sales, customer segments).

Steps:

1️⃣ Select a model

Regression (predict numbers)

Classification (predict categories)

2️⃣ Train the model

Fit the model to your training data to learn patterns.

3️⃣ Predict

Use the model to make predictions on test data.

4️⃣ Evaluate

Measure performance (e.g., accuracy, error rate).

Example (Python):

from sklearn.ensemble import RandomForestRegressor

model = RandomForestRegressor()

model.fit(X\_train, y\_train)

predictions = model.predict(X\_test

TESTING MODEL & DATA PREDICTION:

Testing the model

Here we have tested with the Vgg16 Model With the help of the predict () function.

Steps:

1️⃣ Use test data you set aside during splitting.

2️⃣ Predict outcomes with the trained model:

predictions = model.predict(X\_test)

3️⃣ Evaluate performance:

Regression: Mean Squared Error (MSE), R²

Classification: Accuracy, Precision, Recall

Example (Python):

from sklearn.metrics import mean\_squared\_error

mse = mean\_squared\_error(y\_test, predictions)

print("MSE:", mse)

---

✅ Data Prediction

Goal: Make predictions on new or future data.

Steps:

1️⃣ Prepare new input data (same format as training data).

2️⃣ Use the model:

future\_predictions = model.predict(new\_data)

3️⃣ Save predictions (e.g., CSV) for Tableau visualization or reports.

**SAVING THE MODEL**

Finally, we have chosen the best model now saving that model:

import joblib

joblib.dump(model, "model.pkl")

Load later:

model = joblib.load("model.pkl")

# Objective:

# To analyze cosmetics trends and customer insights using data visualization in Tableau and predict sales or customer behavior with machine learning models.

# Scenarios of Application:

# 1️⃣ Sales Forecasting

# Predict future sales of cosmetics products.

# 2️⃣ Customer Segmentation

# Identify customer groups by preferences and buying behavior.

# 3️⃣ Product Recommendation

# Suggest products to customers based on past purchases.

# 4️⃣ Market Trend Analysis

# Spot emerging trends in cosmetics demand.

# 5️⃣ Promotion Impact

# Simulate how discounts affect sales.

# Tools and Technologies Used:

Python – Data cleaning, modeling, prediction

Scikit-learn – Machine learning algorithms

Pandas – Data manipulation

Tableau – Data visualization and dashboards

Joblib – Model saving and loading

# Concepts and Prerequisites:

# Basic Statistics – Mean, variance, trends

# Machine Learning Basics – Training, testing, evaluation

# Data Preparation – Cleaning and splitting data

# Regression & Classification – Predicting numbers or categories

# Data Visualization – Using Tableau dashboards

# Python Programming – For data handling and modeling

# Methodology:

# Gather cosmetics sales and customer data.

# 2️⃣ Data Preparation

# Clean, transform, and split data into training and testing sets.

# 3️⃣ Model Building

# Train machine learning models to predict trends or behavior.

# 4️⃣ Model Testing & Evaluation

# Assess accuracy using test data.

# 5️⃣ Prediction

# Generate forecasts or classifications on new data.

# 6️⃣ Visualization

# Use Tableau to display insights and

# System Architecture:

# 1 Data Layer

# Collect data (sales, customer info, product details).

# Store in databases or CSV files.

# 2 Processing Layer

# Use Python (Pandas, Scikit-learn) for cleaning, splitting, and modeling.

# Save trained models (Joblib).

# 3. Prediction Layer

# Load models to predict new data.

# 4. Visualization Layer

# Import results into Tableau.

# Create dashboards and reports.

**Application Building:**

**1️⃣ Develop Data Pipeline**

**Collect, clean, and prepare cosmetics data.**

**2️⃣ Build Predictive Models**

**Train models in Python (e.g., sales forecasting).**

**3️⃣ Save and Load Models**

**Use Joblib to store models for reuse.**

**4️⃣ Create Visual Dashboards**

**Import predictions into Tableau for interactive analysis.**

**5️⃣ Deploy and Share**

**Share Tableau dashboards with stakeholders.**

# Results and Observations:

# The model accurately predicted cosmetics sales trends with low error.

# Customer segments with higher purchase likelihood were identified.

# Visual dashboards in Tableau clearly showed seasonal demand peaks.

# Data augmentation improved prediction consistency.

# Conclusion:

# The project successfully analyzed cosmetics trends and predicted sales using machine learning. Tableau visualizations provided clear insights for better decision-making, helping identify customer patterns and forecast demand effectively.

# Future Scope:

Integrate real-time data feeds for live dashboards.

Use advanced models (e.g., deep learning) for better predictions.

Expand to more product categories and regions.

Develop a web app for interactive analysis and reporting

1. **References:**

* Scikit-learn Documentation: https://scikit-learn.org
* Tableau Official Site: https://www.tableau.com
* Pandas Documentation: https://pandas.pydata.org
* Joblib Documentation: https://joblib.readthedocs.io