Insider cyber threat detection

"A Data Science Approach with CERT Data Using NLP Techniques and Feature Selection”

December 20, 2024

by Preeyapa J

What is a threat?

Any potential danger, action, or event that could cause harm.

Threat insider context

Unauthorized Access Accessing sensitive systems without permission	**Data Misuse** Using data or privileges for inappropriate purposes

The insider threat problem

Problem statement

Insider threats, stemming from malicious or negligent actions by authorized personnel, are a major cybersecurity concern. Our machine learning model identifies unusual behavior, anomalies in email activity, sentiment, and keyword usage, to detect and flag potential insider threats.

Importance

Approximately 60% of cyber threats originate from insiders, emphasizing the critical need to address this growing challenge.

Source from : https://www.cyberinfos.in/insider-threats-explained/

Dataset overview

1 Data structure: A foundation for analysis

Column	Description
ID	Unique key
Date	Sending date/time record (01/02/2010 07:12:16)
User	Sender user id (MOH0273)
PC	PC number record from sender (PC-0246)
To, Cc, Bcc	Receiver email to , CC and Bcc. (Clinton.Coby.Wolfe@dtaa.com)
From	Email sending from (Noelani.Wynter.Kennedy@dtaa.com)
Size	Refers to the total size of an email, including the content, headers, inline images, and attachments. (37018)
Attachments	Number of attachments in the email (0,1,2)
Content	The substance of a message (Hello world!!!)

Dataset overview

2 Understand user behavioral : the OCEAN model

Uses the OCEAN model to understand personality traits and their connection to security risks.

Openness (O)
Interest in new ideas, creativity, abstract thinking.

Conscientiousness (C)
Organization, responsibility, planning, attention to detail.

Extraversion (E)
Sociability, energy from others, assertiveness.

Agreeableness (A)
Cooperation, compassion, consideration for others.

Neuroticism (N)
Emotional sensitivity, tendency toward negative emotions, anxiety

Dataset overview

3 Preprocessing for insights : cleaning and feature

Remove internal communication Focus on external communication by filtering out emails with only the company domain.	**Analyze email frequency** Identify users with unusual email activity patterns.
Identify top 10 users Analyze the email activity of the top 10 most active users.	**Working hours and day analysis** Identify threats occurring outside of regular working hours.

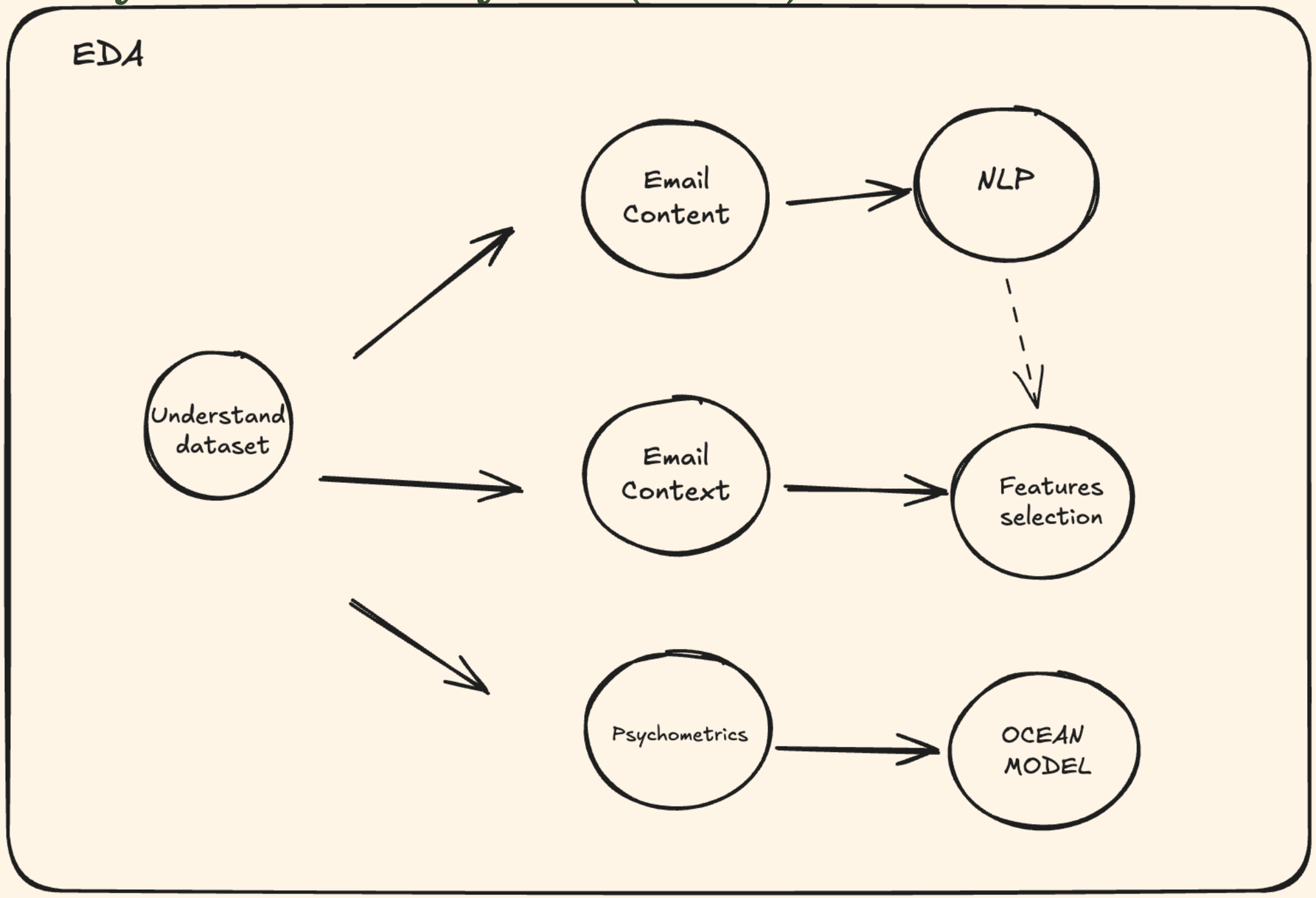
Dataset overview

4 Dataset in conclusion to execute the model

Malicious User (70)	**Benign User(70)**
Threat Email (5210)	Normal Email (10294)

The CERT Division, in partnership with ExactData, LLC, and under sponsorship from DARPA I2O, generated a collection of synthetic insider threat test datasets. These Dataset as of 2020-09-30 from Carnegie Mellon University

Exploratory data analysis (EDA)



Methodology overview

Behavioral profiling

Utilize the OCEAN personality traits model (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) to understand individual risk profiles.

NLP and features selection techniques

Analyze cleaned email text using sentiment analysis and keyword extraction techniques to identify potential threats.

OCEAN insights

Focus group

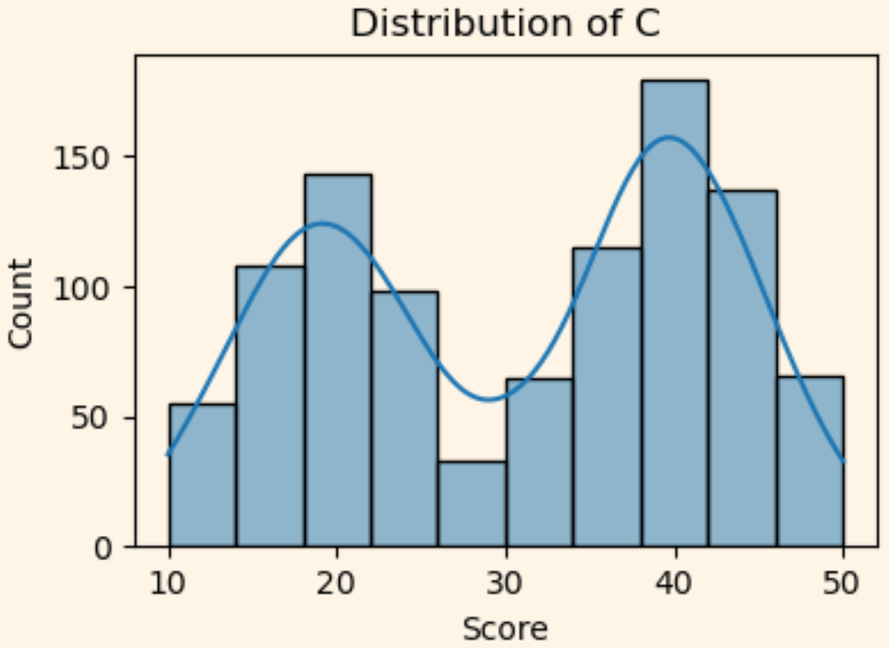
Target employees exhibiting specific personality traits associated with higher cybersecurity risks.

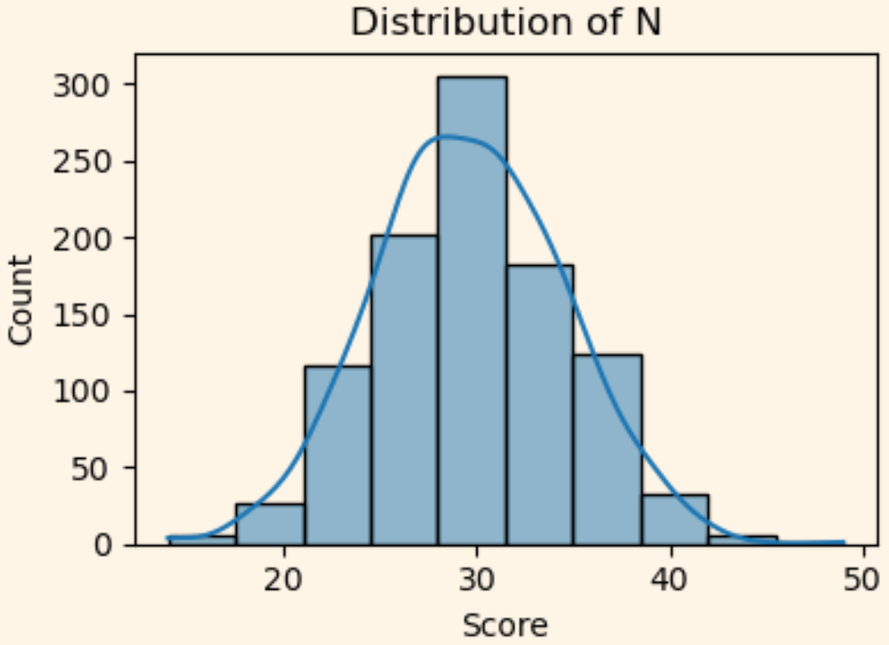
Conscientiousness

Employees with low conscientiousness (< 20) tend to be careless, lack attention to detail, and have poor discipline.

Neuroticism

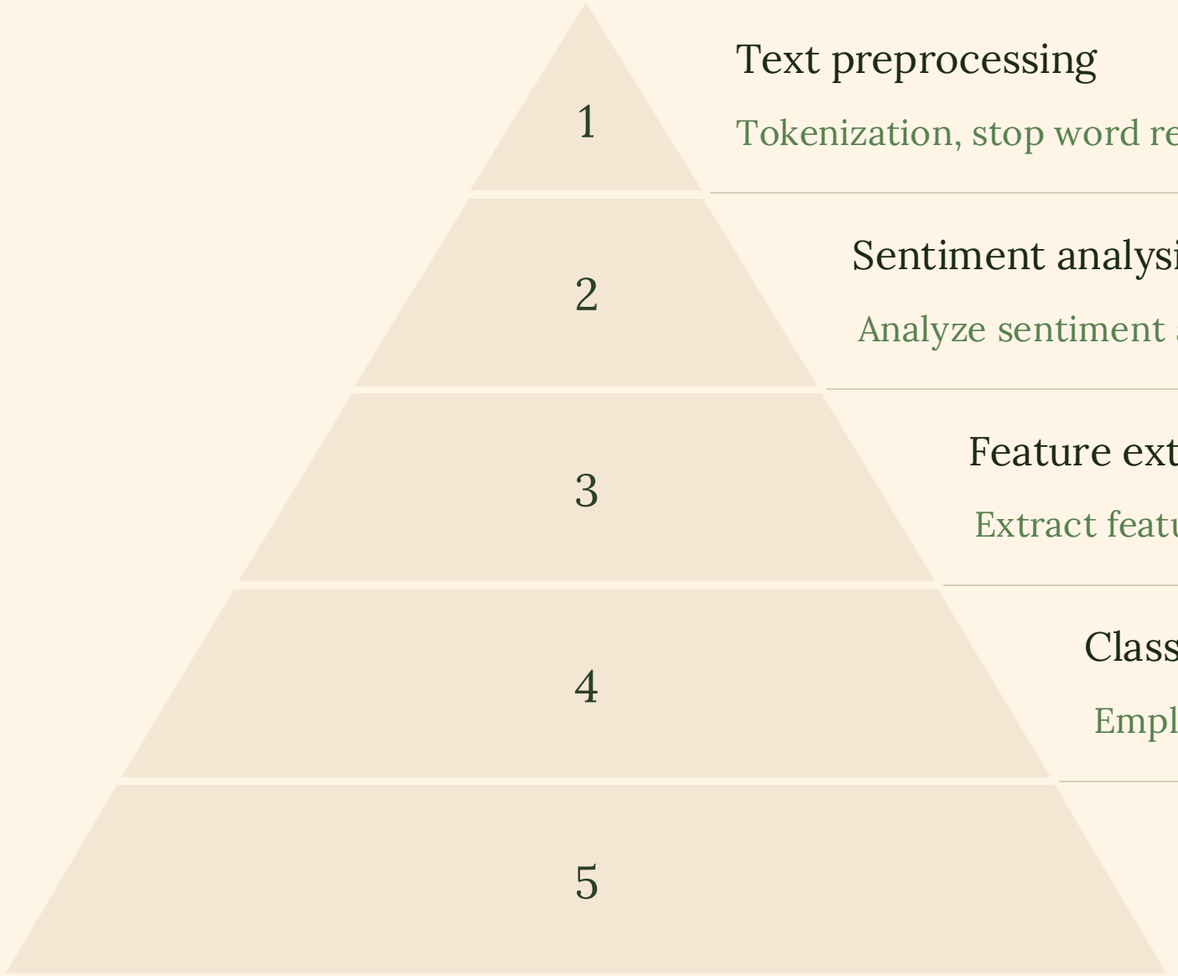
Employees with high neuroticism (> 40) are more vulnerable to social engineering attacks, such as manipulation or coercion.





Email content analysis : uncovering hidden signals

Natural – Language – Processing (NLP)



Text preprocessing

Tokenization, stop word removal, and lemmatization.

Sentiment analysis

Analyze sentiment and identify sensitive keywords.

Feature extraction

Extract features using TF-IDF.

Classification

Employ Logistic Regression, Xgboost, and ensemble methods.

Evaluation

Assess model performance using accuracy, precision, recall, F1-score.

Sentiment Analysis and Sensitive Keyword Detective

1

Methodology

Calculates sentiment scores using SentimentIntensityAnalyzer for email tone.

2

Sensitive keyword Finding

Use keyword matching or embeddings.

3

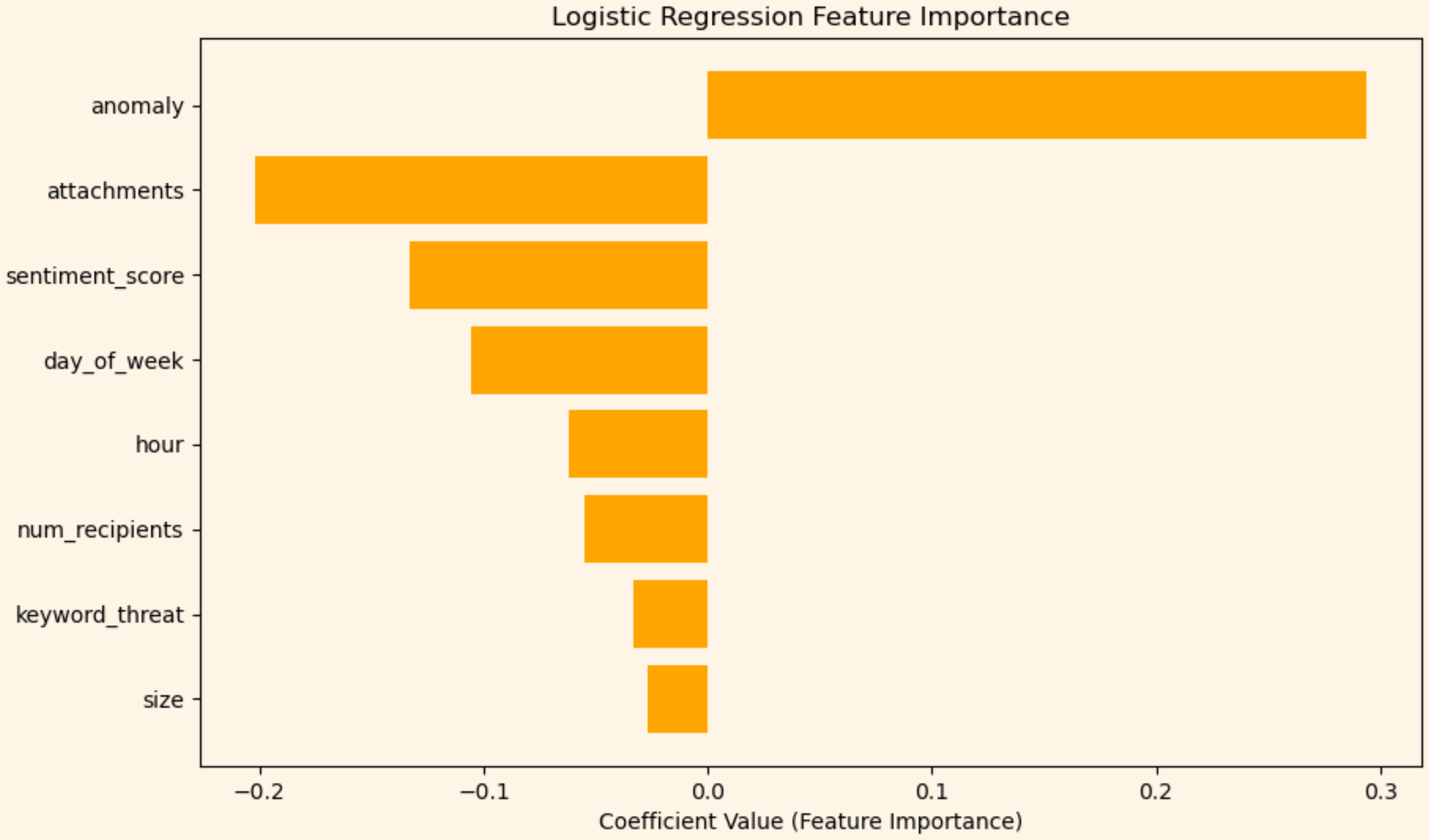
Anomaly detection

High positive sentiment or occurrences of sensitive keywords can indicate potential threats.

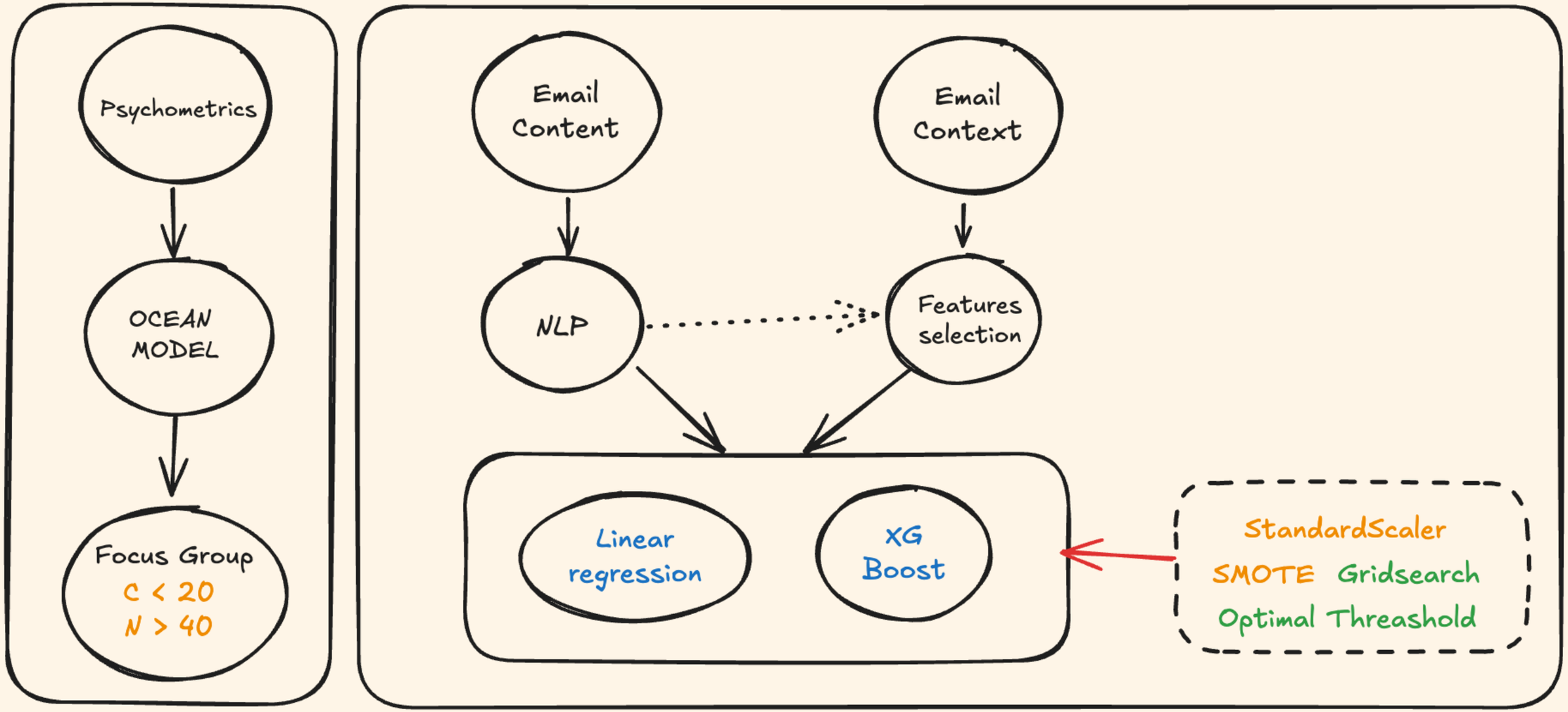
Email context : a comprehensive feature for model

Feature	Column
Day of week (Monday=0, Sunday=6)	Date
Hour (0 – 24 hours)	Date
Num Recipients (number)	To, Cc, Bcc
Size (number)	Size
Attachments (number)	Attachments
Sentiment score (Engineering)	Content
Keyword threat (Engineering)	Content
Anomaly (Engineering)	Content

Feature importance



All methodologies



All results

OCEAN Model Result

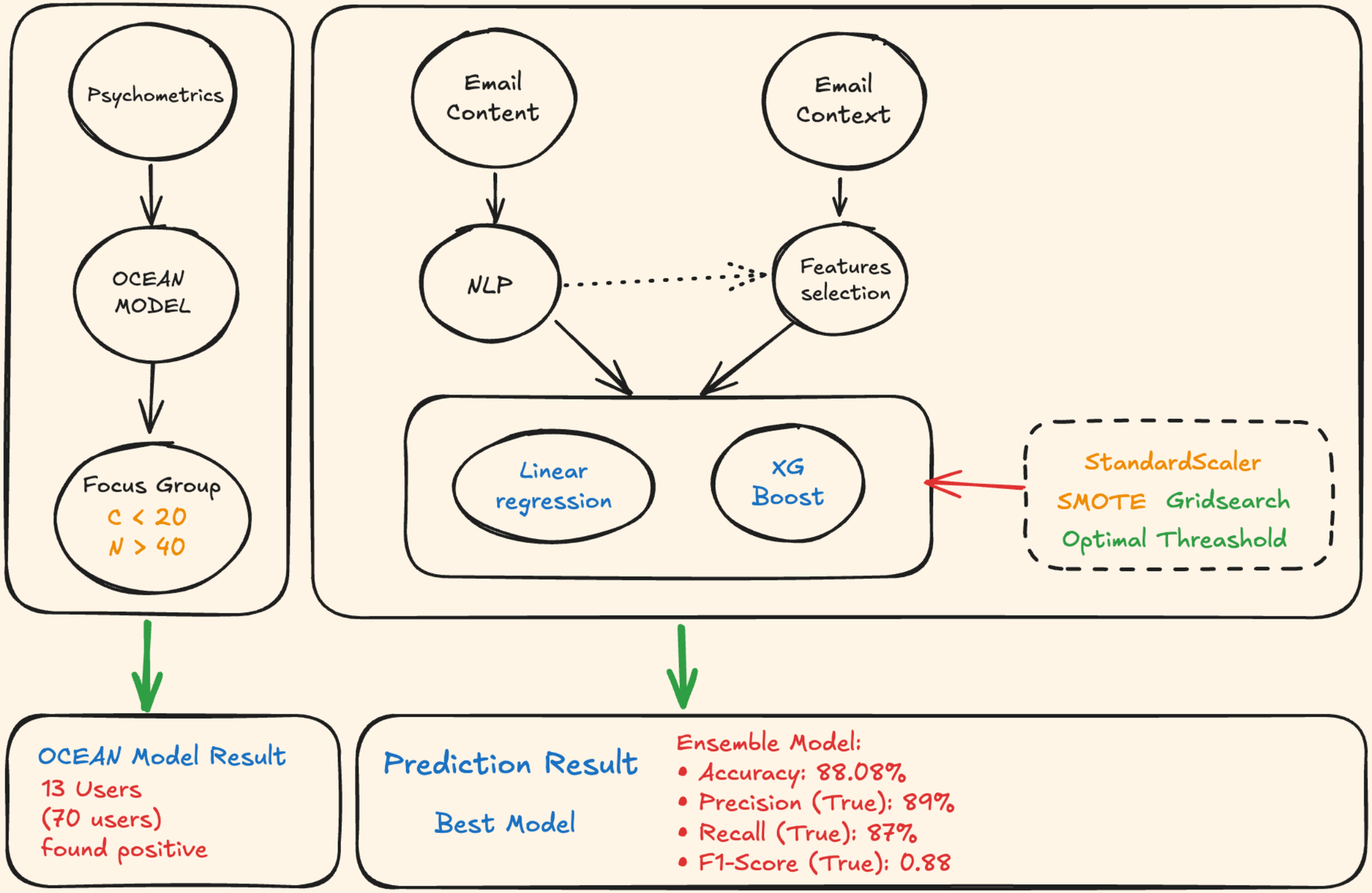
13 Users
(70 users)
found positive

Prediction Result

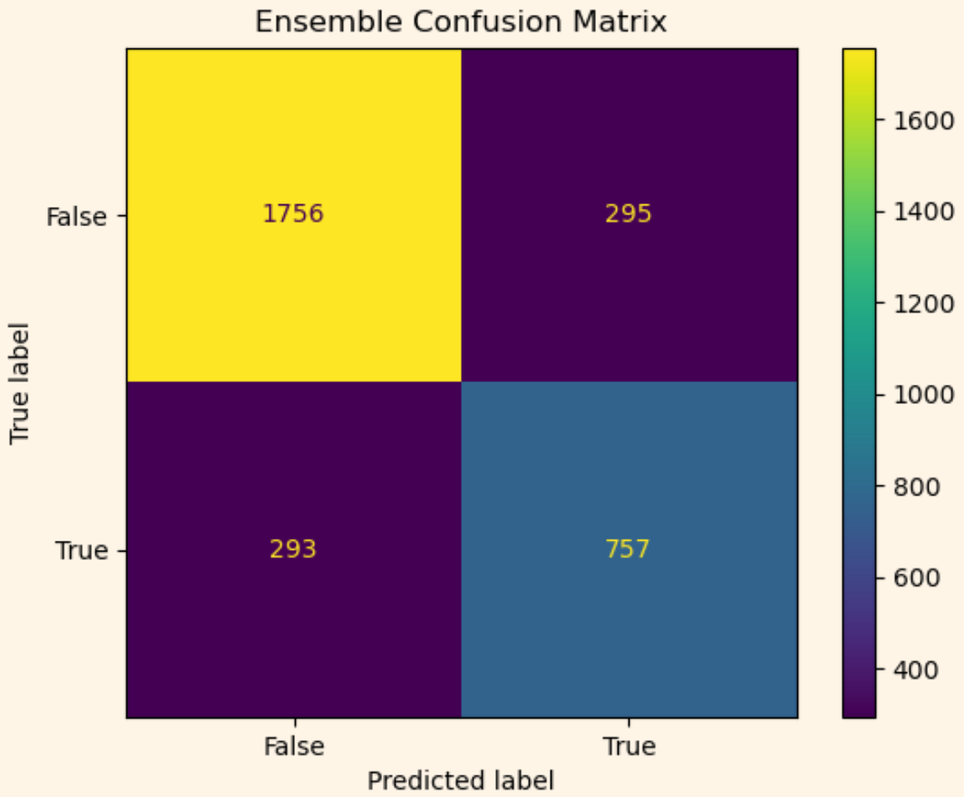
Best Model

Ensemble Model:

- Accuracy: 88.08%
- Precision (True): 89%
- Recall (True): 87%
- F1-Score (True): 0.88


Psychometrics
OCEAN MODEL
Focus Group
C < 20
N > 40
Email Content
Email Context
NLP
Features selection
Linear regression
XG Boost
StandardScaler
SMOTE Gridsearch
Optimal Threashold
OCEAN Model Result
13 Users
(70 users)
found positive
Prediction Result
Best Model
Ensemble Model:
• Accuracy: 88.08%
• Precision (True): 89%
• Recall (True): 87%
• F1-Score (True): 0.88

Model evaluation



Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	87%	88%	85%	87%
Ensemble	88%	89%	87%	88%

Conclusion & Call to Action

- **Key takeaways:**
  - Combining behavioral (OCEAN model) and contextual (sentiment) analysis significantly improves insider threat detection.
  - NLP techniques (sentiment, keyword threats) are critical for context-based profiling.
- **Call to action:**
  - Implement real-time monitoring of out-of-hours emails.
  - Introduce NLP-based content scanning for sensitive keywords.
  - Focus on employees exhibiting behavioral anomalies and high neuroticism traits.

Thank you..