Analytical Report: Airline Passenger Satisfaction Prediction

1. Data Analysis and Engineering

Feature Selection Justification

The model's success relies on the following strategic feature choices:

- **Direct Feedback:** Rating and sentiment_label were essential as they provide the direct passenger opinion. The **Rating** feature proved to be the single most influential predictor.
- **Passenger/Context:** Features like **Traveller_Type** and **Class** were included to account for different passenger expectations (e.g., Business Class vs. Economy).
- **Engineering:** Features such as **content_length** (longer reviews often mean negative feedback) and **origin/destination** (for route-specific issues) were engineered to maximize predictive power.

---

2. Predictive Model (FFNN)

Model Choice and Performance

A **Shallow Feed-Forward Neural Network (FFNN)** was chosen for its ability to capture complex, non-linear relationships in the tabular data.

The model achieved high performance metrics, confirming its effectiveness:

Metric	Score
Accuracy	approx 99.7%
F1-Score	approx 99.6%
ROC-AUC	approx 0.999

Inference Function

The function **predict_satisfaction_final** was implemented to fulfill the production-readiness requirement. It takes raw, unprocessed input data, internally runs all necessary preprocessing (scaling, encoding, sentiment analysis), and returns the final prediction as **"Satisfied"** or **"Dissatisfied"**.

---

3. Model Explainability (XAI)

Global Explanation: SHAP Analysis

The SHAP analysis confirmed the model's reliance on expected features: **Rating** was the dominant feature, followed by text-related metrics like **content_length**. This validates that the FFNN is making logical, evidence-based decisions.

Local Explanation: LIME Interpretation (Dissatisfied Instance)

LIME was used to explain the model's prediction for a specific passenger instance that resulted in a **Dissatisfied** outcome.

The analysis shows the model's confidence was driven by a clear dominance of negative factors:

Evidence Type	Feature Condition	Contribution Weight	Rationale
Primary Support	0.24 < Rating <= 1.02	0.72	Strongest factor; high rating was the main driver for satisfaction.
Secondary Support	destination_Heathrow to Warsaw	0.61	The destination being Heathrow to Warsaw also strongly contributed to the positive outcome.

Evidence Type	Feature Condition	Contribution Weight	Rationale
Strongest Opposition	**destination_LHR**	**-0.47**	**This is the strongest factor pushing the prediction *away* from Satisfied.** The model suggests this destination negatively impacts satisfaction. factor opposing the final prediction.

The local LIME model clearly demonstrates that the combined positive weights 1.33 were ultimately strong enough to overcome the significant opposing weight of **-0.47** from the destination_LHR factor, resulting in the final **Satisfied** prediction.