THE MACHINE LEARNING CANVAS

Designed for: Designed by: Date: Iteration: .

PREDICTION TASK

Type of task? Entity on which predictions are made? Possible outcomes? Wait time before observation?

Type of Task:

Classification Task

Entity on Which Predictions Are Made:

Flights or Carrier Operations

Possible Outcomes:

On-time (1)

Delayed (0)

Wait Time Before Observation:

Post-event Data Availability: The predictions would be made after the data for the entire observation period (e.g., a specific time range in 2022) has been collected and processed. This means the wait time before observation would be the time it takes to collect, clean, and prepare the dataset for analysis.

DECISIONS

How are predictions turned into proposed value for the end-user? Mention parameters of the process / application that does that.

Our project aims to enhance the on-time performance predictions of carriers using machine learning models trained on Marketing Carrier On-Time Performance data from the TranStats database. The feasibility of this project is supported by a robust inventory of resources including skilled personnel, diverse data sources, advanced computing resources, and a comprehensive suite of software tools.

Key Steps to Turn Predictions into Proposed Value:

1. **Prediction Interpretation and Contextualization:**
    - Upon generating predictions using our machine learning models, it is crucial to interpret these predictions in a manner that is understandable and actionable for end-users. For instance, predictions might indicate the likelihood of on-time performance for different carriers or the probability of delays.
2. **Integration with User Interfaces and Decision-Making Tools:**
    - Delivering predictions effectively involves integrating them into existing user interfaces, dashboards, or

VALUE PROPOSITION

Who is the end-user? What are their objectives? How will they benefit from the ML system? Mention workflow/interfaces.

End-User:

The end-users of the ML system are primarily the operations team and management at an airline or carrier company.

Objectives:

- **Operations Team:**
    - Minimize flight delays.
    - Optimize scheduling and resource allocation.
    - Enhance customer satisfaction by ensuring timely flights.
- **Management:**
    - Improve overall operational efficiency.
    - Reduce operational costs.
    - Increase profitability and competitive advantage.

Benefits from the ML System:

- **Improved Decision-Making:** The ML system provides accurate predictions about flight delays, enabling the operations team to make proactive decisions.
- **Operational Efficiency:** By anticipating delays, the system helps in optimizing

DATA COLLECTION

Strategy for initial train set & continuous update. Mention collection rate, holdout on production entities, cost/constraints to observe outcomes.

The objective of this strategy is to outline how we will manage the initial training set and continuously update our machine learning models to ensure accuracy and relevance over time, considering collection rates, holdouts on production entities, and cost/constraints to observe outcomes effectively.

Initial Train Set Strategy:

1. **Data Collection and Preparation:**
    - **Data Sources:** Utilize Marketing Carrier On-Time Performance data from the TranStats database for initial model training.
    - **Data Extraction:** Ensure comprehensive extraction of historical data to capture variability and trends in carrier performance.
    - **Data Cleaning and Integration:** Apply robust data cleaning techniques to handle missing values and outliers, ensuring data quality before training.
2. **Feature Engineering:**
    - **Feature Selection:** Identify relevant features (e.g., flight schedules, weather conditions) that impact carrier on-time performance.
    - **Feature Transformation:** Transform raw data into meaningful features that enhance model prediction accuracy.
3. **Model Selection and Training:**
    - **Algorithm Choice:** Select appropriate algorithms (e.g.,

DATA SOURCES

Where can we get (raw) information on entities and observed outcomes? Mention database tables, API methods, websites to scrape, etc.

https://www.kaggle.com/datasets/robikscube/flight-delay-dataset-20182022 decision-making tools that our stakeholders already utilize. This ensures predictions are accessible and seamlessly fit into their workflow.

3. **Presentation of Insights and Recommendations:**
    - Besides presenting predictions, it is essential to provide insights derived from these predictions. This includes explaining the factors influencing predictions and suggesting potential actions or decisions based on the predicted outcomes. For example, recommending adjustments to scheduling or operational planning based on predicted delays.
4. **Value Proposition for Stakeholders:**
    - Articulate clearly how these predictions add value to our stakeholders. Improved accuracy in predicting on-time performance can lead to enhanced operational efficiency, reduced costs associated with delays, and improved customer satisfaction due to better service schedules, re-routing flights, and better resource management.

- **Cost Savings:** Reducing delays and optimizing operations lead to significant cost savings, such as reduced fuel costs and minimized overtime payments.
- **Customer Satisfaction:** Timely flights enhance the passenger experience, leading to higher customer satisfaction and loyalty.

Example Workflow:

Data Input:

1) **Data Source:** The Marketing Carrier On-Time Performance data table for 2022 from the "On-Time" database of the TranStats data library.
2) **Method:** Data is loaded in batches into the data processing pipeline.

Prediction Engine:

1) **Tools Used:** Python, scikit-learn, pandas
2) **Method:** The data is processed and predictions are made using a pre-trained ML model. The model is developed and maintained by Ivan Golov (Data Scientist) and Artem Bulgakov (ML Engineer).

Decision Interface:

Tools Used: FastAPI, Flask

Method: Predictions are displayed on an interactive dashboard. The dashboard shows the likelihood of delays and other insights, allowing users to make informed decisions.

Alerts:

1) **Tools Used:**

regression, classification) based on the nature of the prediction task (e.g., predicting on-time vs. delayed flights).
    - **Model Training:** Train initial models using a subset of historical data, focusing on model interpretability and performance metrics (e.g., accuracy, precision, recall).
4. **Validation and Evaluation:**
    - **Holdout Set:** Reserve a holdout set from initial data to evaluate model performance objectively.
    - **Cross-Validation:** Implement cross-validation techniques to assess model robustness and generalizability.
5. **Deployment Readiness:**
    - **Model Validation:** Validate models against predefined business metrics and thresholds to ensure readiness for deployment.
    - **Documentation:** Document model performance, assumptions, and limitations for future reference and continuous improvement.

Continuous Update Strategy:

1. **Real-Time Data Integration:**
    - **Data Pipeline:** Establish a robust data pipeline using Apache Airflow and Apache Pulsar to ingest real-time data updates.
    - **Streaming Data Processing:** Implement streaming data processing to handle continuous updates efficiently.
2. **Incremental Learning and Retraining:**
    - **Incremental Updates:** Implement mechanisms (e.g., online learning,

reliability.

5. **Continuous Improvement and Feedback Mechanism:**
   - Establish a feedback loop where stakeholders can provide input on the accuracy and usefulness of predictions. This feedback helps refine our models and ensure that the proposed values align with stakeholder expectations and business objectives.

Implementation Considerations:

- **User-Centric Design:** Ensure predictions and insights are presented in a user-friendly manner, considering the varying technical expertise and roles of end-users.
- **Business Integration:** Align predictions with business goals and operational processes to ensure they directly contribute to strategic objectives such as cost reduction and customer satisfaction.
- **Scalability and Sustainability:** Plan for the scalability of prediction delivery mechanisms as the project progresses, ensuring that as more carriers and data are integrated, predictions remain timely and relevant.

Integration with messaging systems (e.g., Slack, email, SMS)

2) **Method:** If a flight is predicted to be delayed, an alert is triggered, notifying the operations team and relevant stakeholders.

Recommendations:

1) **Tools Used:** Custom algorithms implemented in Python, PostgreSQL for storing decision rules
2) **Method:** The system suggests actions to mitigate potential delays. Recommendations are based on historical data and predictive models, helping to optimize operations and improve on-time performance.

Feedback and Retraining:

1) **Tools Used:** MLflow for tracking model performance, Apache Airflow for scheduling retraining
2) **Method:** Model performance is monitored continuously. Feedback is collected, and the model is retrained periodically to ensure it adapts to new data patterns and maintains high accuracy.

By following this workflow, predictions can be effectively transformed into actionable insights, providing significant value to end-users by improving decision-making and operational efficiency.

batch updates) to incorporate new data into existing models without retraining from scratch.
   - **Trigger Mechanisms:** Set triggers based on data drift or model performance degradation to initiate retraining cycles.

3. **Model Monitoring and Evaluation:**
   - **Monitoring Framework:** Deploy monitoring frameworks (e.g., MLflow, ClearML) to track model performance metrics and data quality over time.
   - **Alert Mechanisms:** Set up alert mechanisms to notify stakeholders of anomalies or deviations in model predictions.
4. **Cost and Constraints Consideration:**

- **Collection Rate:** Ensure data collection rates are sufficient to capture real-time changes in carrier performance, balancing the frequency of updates with operational costs.
- **Holdout on Production Entities:** Maintain holdout sets on production entities to validate model performance against current data, ensuring models generalize well to unseen scenarios.
- **Cost and Resource Allocation:** Allocate resources (e.g., computing resources, personnel time) effectively to manage the costs associated with continuous model updates and monitoring.

Conclusion: This strategy outlines a structured approach to managing the initial train set and continuous model updates for predicting carrier on-time performance. By focusing on data quality, model robustness, and continuous monitoring, we aim to maintain high prediction accuracy while adapting to evolving data and business conditions. Continuous evaluation and adherence to cost and constraint considerations will ensure sustainable and effective deployment of machine learning models in our project.

IMPACT SIMULATION

Can models be deployed? Which test data to assess performance? Cost/gain values for (in)correct decisions? Fairness constraint?

Can Models be Deployed?

Yes, the machine learning models developed for predicting carrier on-time performance can be deployed into production. Deployment involves making the models accessible within the operational environment where they can generate predictions in real-time or on-demand basis.

Assessment of Model Performance

1. **Test Data for Performance Assessment:**
   - **Holdout Set:** Use a holdout set of production data that was not

MAKING PREDICTIONS

When do we make real-time / batch pred.? Time available for this + featurization + post-processing? Compute target?

1. Real-Time vs. Batch Predictions:

- **Real-Time Predictions:** These predictions are generated on-the-fly as new data becomes available. For predicting carrier on-time performance, real-time predictions would involve processing data immediately upon receipt, such as flight schedule updates or weather conditions.
- **Batch Predictions:** Batch predictions involve processing a large volume of data in batches, typically at scheduled intervals. This approach is useful for handling historical data or large datasets where real-time processing may not be necessary.

2. Featurization:

- **Featurization Process:** This step involves transforming raw data into features that can be used by the machine learning models for prediction. Featurization includes extracting relevant information from data sources (e.g., flight schedules, weather data), performing calculations or aggregations, and preparing the data in a format suitable for model input.
- **Time Required:** Featurization time depends on the complexity and volume of data. Real-time featurization aims for minimal latency to ensure

BUILDING MODELS

How many prod models are needed? When would we update? Time available for this (including **featurization and analysis)?**

Number of Production Models:

The number of production models needed for predicting carrier on-time performance depends on several factors:

- Use Case Complexity: If different aspects of on-time performance (e.g., different carriers, routes, weather conditions) require distinct modeling approaches, multiple models might be necessary.
- Granularity: Consider the granularity of predictions needed. For example, a single model might suffice for predicting overall on-time performance across all carriers, or separate models might be required for individual carriers or routes.
- Specialized Requirements: Specific requirements such as fairness constraints (ensuring predictions are unbiased across different groups) or regional variations may necessitate multiple models tailored to different contexts.

2. Update Frequency:
The update frequency of production models involves:

- Data Freshness: Models should be updated when new data becomes available to maintain accuracy and relevance. For instance, updating models daily or weekly based on the frequency of data updates (e.g., flight schedules, weather forecasts).
- Performance Monitoring: Monitor model performance metrics continuously. Update models when performance degrades beyond

FEATURES

Input representations available at prediction time, extracted from raw data sources.

At prediction time, the input representations for predicting carrier on-time performance can be derived from various raw data sources. These features are crucial as they directly influence the accuracy and reliability of the predictions. Here's an outline of the input representations typically available:

1. **Flight Schedule Information:**
   - **Departure and Arrival Times:** Scheduled departure and arrival times of flights.
   - **Flight Duration:** Estimated duration of the flight.
   - **Route**

used during model training to assess performance. This ensures that the model is evaluated on unseen data, providing a realistic measure of its generalization capability.

- **Real-time Monitoring:** Continuously monitor model performance using incoming real-time data to validate predictions and detect any degradation in accuracy

timely predictions, while batch featurization can leverage more computation resources for thorough data preparation.

3. Post-Processing:

- **Post-Processing Steps:** After predictions are generated, post-processing involves interpreting the model outputs, applying business rules or thresholds, and formatting results for user consumption or further decision-making.
- **Time Considerations:** Post-processing time varies based on the complexity of the business logic applied and the format of the output required. Real-time systems prioritize low-latency responses, while batch processes allow for more extensive analysis and reporting.

4. Computing Target:

- **Target System:** Determine the computing infrastructure required for making predictions based on the volume and velocity of incoming data, the complexity of models, and the latency requirements.
- **Scalability:** Ensure the computing target is scalable to handle fluctuations in data volume and processing demands, especially during peak times or sudden changes in operational conditions.

Time Available for Prediction Processes:

- **Real-Time Predictions:** Typically, real-time predictions aim for low-latency processing, often requiring predictions within milliseconds to seconds depending on the application (e.g., real-time

acceptable thresholds or when new insights from data analysis suggest improvements.

- Business Needs: Align update frequency with business needs and operational changes. For example, during peak travel seasons or significant operational changes, more frequent updates may be necessary.

3. Time Available for Updates (Including Featurization and Analysis):

- Featurization Time: Featurization time depends on the complexity of data transformations required for model input. Real-time featurization aims for minimal latency, while batch featurization may take longer but allows for more comprehensive data processing.
- Analysis Time: After featurization, allocate time for model training, evaluation, and validation using updated data. This includes testing new features, tuning hyperparameters, and assessing model performance against established metrics.
- Deployment Time: Time to deploy updated models into production systems, ensuring seamless integration and minimal disruption to operational processes.

Example Scenario:

- Number of Models: Initially, start with a single model predicting overall carrier on-time performance. As operational insights and data granularity increase, consider deploying additional models for specific carriers or regions if performance gains justify the added complexity.
- Update Frequency: Update models weekly to incorporate new data on flight schedules, weather

Information: Origin and destination airports, flight numbers, and airline carriers.

2. **Weather Conditions:**
   - **Current Weather:** Weather conditions at departure and arrival airports.
   - **Forecasted Weather:** Predicted weather conditions during the flight duration.
   - **Weather Alerts:** Alerts for severe weather conditions that could impact flight operations

over time.

2. **Cost/Gain Values for (In)correct Decisions:**
   - **Cost of Incorrect Decisions:** Evaluate the potential costs associated with incorrect predictions (e.g., financial penalties for misjudged flight delays, customer dissatisfaction).
   - **Gain from Correct Decisions:** Estimate the benefits derived from accurate predictions (e.g., improved

monitoring of flight status).

- **Batch Predictions:** Batch prediction processes are scheduled based on operational needs and available computing resources. These processes may take minutes to hours, depending on the size of the dataset and complexity of featurization and model inference.

Conclusion:

Balancing real-time and batch prediction methods involves optimizing featurization, post-processing, and computing resources to meet operational requirements. Understanding the time constraints and computing targets ensures efficient deployment and utilization of machine learning models for predicting carrier on-time performance.

conditions, and operational changes. Featurization and analysis processes should be streamlined to fit within the update timeframe, ensuring models reflect the most current information available.

Conclusion:
Determining the number of production models and their update frequency involves balancing complexity, operational needs, and data availability. By aligning model updates with business objectives and leveraging efficient featurization and analysis processes, organizations can maintain high-performance predictive capabilities in dynamic operational environments.

.

3. **Historical Performance Data:**
   - **Historical On-Time Performance:** Past records of on-time performance for the carrier and specific routes.
   - **Delay Patterns:** Patterns of delays due to weather, congestion, or operational issues.
4. **Operational Data:**
   - **Aircraft Type:** Type of aircraft operating the flight.
   - **Gate and Terminal Information:**

operational efficiency, enhanced customer experience). 3. **Fairness Constraint:** ○ **Fairness Evaluation:** Assess model fairness to ensure predictions do not exhibit bias towards particular groups (e.g., airlines, demographic segments). Use fairness metrics and techniques to detect and mitigate biases in predictions. ○ **Fairness**				Specific gate and terminal assignments. ○ **Air Traffic Control (ATC) Delays:** Delays caused by air traffic control. 5. **External Factors:** ○ **Economic Indicators:** Factors such as fuel prices or economic conditions influencing airline operations. ○ **Political Events:** Events affecting airspace regulations or airport operations.

Monitoring: Implement ongoing monitoring of model predictions to detect and address any fairness issues that may arise as new data is processed. **Cost/Gain Values Example:** • **Cost of Incorrect Decision:** A false negative (predicting a flight will be on-time when it will actually be delayed) could lead to increased operational costs due to disrupted schedules and potential compensation to passengers. • **Gain from**				6. **Real-Time Updates:** ○ **Current Flight Status:** Real-time updates on flight delays, gate changes, or cancellations. ○ **Weather Updates:** Updated weather conditions affecting flight paths. **Example Input Representation:** For predicting whether a flight will be on-time or delayed, the input features might include: • Scheduled departure time. • Scheduled arrival time. • Current weather conditions at departure

Correct Decision: A true positive (accurately predicting a flight will be delayed) allows airlines to proactively manage resources, potentially reducing costs associated with last-minute adjustments and improving overall customer satisfaction by providing timely information. **Fairness Constraint Example:** ● **Fairness Assessment:** Use metrics such as demographic parity or equal opportunity to ensure predictions do not unfairly disadvantage certain carriers or passenger groups. ● **Fairness**				airport. ● Forecasted weather conditions at arrival airport. ● Historical on-time performance of the carrier on the specific route. ● Aircraft type and capacity. ● Air traffic control delays expected during the flight duration. **Integration and Featurization:** These raw data sources undergo integration and featurization processes before being used for prediction: ● **Integration:** Data from various sources (e.g., weather APIs, airline databases) are integrated into a unified data pipeline.

Monitoring: Implement ongoing monitoring and auditing processes to detect biases in predictions and take corrective actions as needed, ensuring fairness in model outcomes. **Conclusion:** Deploying machine learning models for predicting carrier on-time performance is feasible, provided rigorous assessment of model performance using appropriate test data, consideration of cost and gain implications of predictions, and adherence to fairness constraints to ensure equitable outcomes. Continuous monitoring and refinement of models are essential to maintain accuracy, fairness, and relevance in dynamic operational environments.				• **Featurization:** Raw data is processed to extract relevant features, handle missing values, and transform data into a format suitable for machine learning models. By leveraging these input representations effectively at prediction time, machine learning models can generate accurate predictions of carrier on-time performance, contributing to improved operational efficiency and customer satisfaction in the airline industry.

MONITORING

Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)?

Monitoring the impact and value creation of a machine learning (ML) system in production involves tracking specific metrics that assess both technical performance and business outcomes. Here are key metrics to quantify value creation and measure the ML system's impact on end-users and the business:

Technical Performance Metrics:

1. **Prediction Accuracy:**
   - **Definition:** The percentage of correct predictions made by the model compared to actual outcomes.
   - **Importance:** Ensures the model's reliability in generating accurate predictions of carrier on-time performance.
2. **Precision and Recall:**
   - **Precision:** Proportion of correctly predicted on-time flights among all flights predicted to be on-time.
   - **Recall:** Proportion of correctly predicted on-time flights among all actual on-time flights.
   - **Importance:** Provides insights into the model's ability to

minimize false positives (incorrectly predicting delays) and false negatives (missing actual delays).

3. **F1 Score:**
    - **Definition:** Harmonic mean of precision and recall, balancing between precision and recall.
    - **Importance:** Offers a single metric to evaluate overall model performance, especially useful when precision and recall need to be balanced.

Business Impact Metrics:

1. **Cost Savings:**
    - **Definition:** Financial savings achieved through improved operational efficiency (e.g., reduced costs associated with flight delays, optimized resource allocation).
    - **Importance:** Quantifies direct monetary benefits derived from accurate predictions and proactive decision-making.
2. **Customer Satisfaction Metrics:**
    - **Customer Complaints:** Number of customer complaints related to flight

delays or on-time performance.

- **Net Promoter Score (NPS):** Measurement of customer satisfaction and loyalty based on customer surveys.
- **Importance:** Reflects how well the ML system's predictions align with customer expectations and overall satisfaction levels.

3. **Operational Efficiency:**
   - **On-Time Performance Improvement:** Percentage increase in on-time flights compared to historical averages.
   - **Resource Optimization:** Efficiency gains in resource allocation (e.g., crew scheduling, gate assignments).
   - **Importance:** Demonstrates operational benefits achieved through timely and accurate predictions.

System Health and Monitoring Metrics:

1. **Model Performance Metrics:**
   - **Model Accuracy Over Time:** Trend of prediction accuracy as new data is incorporated and models are updated.
   - **Drift Detection:** Monitoring changes in data

distribution or model performance that may indicate the need for model retraining.
    - **Importance:** Ensures models remain effective and reliable over the long term, adapting to evolving data patterns and operational conditions.
2. **Latency and Throughput:**
    - **Prediction Latency:** Time taken to generate predictions from input data.
    - **Prediction Throughput:** Number of predictions processed per unit time.
    - **Importance:** Measures the responsiveness and scalability of the ML system, ensuring predictions are generated in a timely manner to support operational decision-making.

Example Scenario:

- **Metric:** Percentage Increase in On-Time Performance (Business Impact)
    - **Definition:** Measure the percentage improvement in the number of flights arriving on-time compared to historical performance.
    - **Importance:** Directly correlates to improved customer

satisfaction, reduced operational costs, and enhanced overall business efficiency.

Conclusion:

Monitoring these metrics provides a comprehensive view of the ML system's impact on both technical performance and business outcomes. Regularly evaluating these metrics enables stakeholders to make data-driven decisions, optimize model performance, and maximize the value created by the ML system in predicting carrier on-time performance.