CALIFORNIA STATE UNIVERSITY
LONG BEACH
College of Business

IS 675: Deep Learning for Business

Professor: Basil Latif

Transport analytics: Flights and Ports

TEAM 7

By:

Antariksh Ramesh – 032175817

Brunoo Medaballimi Bosco – 032171462

Prithviraju Venkataraman – 032163597

Rishitha Kandibyalla – 032216832

Table of Contents

Abstract

Transportation systems, whether in the skies or across the seas, are the lifelines of global trade. They keep goods moving and connect people and businesses worldwide. However, these systems are far from perfect—delays and inefficiencies are common, causing disruptions in supply chains, increasing costs, and frustrating both businesses and customers. Understanding and addressing these challenges is crucial to keeping the wheels of trade turning smoothly.

This project focuses on tackling these issues in three main areas: predicting flight delays, analyzing port operations, and identifying congestion using aerial imagery. By combining different types of data—like flight schedules, weather conditions, port cargo volumes, and aerial images of planes and ships—we use advanced machine learning (ML) and deep learning (DL) techniques to uncover patterns and propose solutions.

Through predictive models, we identify the key factors behind flight delays, offering insights that can help airlines optimize their schedules and reduce disruptions. In ports, our analysis highlights trends in cargo movement and congestion, helping managers allocate resources more effectively during peak periods. Using state-of-the-art image recognition techniques, we also analyze aerial images to detect real-time congestion in airports and seaports, giving a visual perspective that complements the data analysis.

By bringing together structured data and visual insights, this project provides a complete picture of the inefficiencies in transportation systems. More importantly, it delivers practical recommendations that can make real-world operations smoother, more efficient, and more reliable. Whether it's helping airlines reduce delays or improving the flow of goods at ports, this study demonstrates how technology and data can transform the way we manage transportation, making it ready to meet the challenges of today and tomorrow.

Introduction

Transportation networks are the backbone of the modern economy, ensuring the smooth movement of goods and people across the globe. These networks connect industries, enable trade, and support millions of daily activities. Airports and seaports, in particular, serve as critical nodes in these networks, acting as hubs where vast amounts of goods and passengers pass through every day. The efficiency of these hubs directly influences global supply chains, business operations, and economic stability. However, despite their importance, these systems face significant challenges that can create widespread disruptions.

A single delayed flight doesn't just inconvenience passengers. It can trigger a domino effect of missed connections, rescheduling headaches, and added operational costs for airlines. Flight crews may miss their next assignments, aircraft turnaround times may increase, and passengers may experience delays in reaching their destinations. These disruptions can extend far beyond the original delay, affecting subsequent flights and entire airline networks.

Similarly, congestion at ports during busy trade seasons can lead to cargo sitting idle for days or even weeks. These impacts businesses relying on timely deliveries, disrupts manufacturing schedules, and increases the cost of holding inventory. Delays in port operations ripple through supply chains, affecting everything from retail availability to industrial production. These inefficiencies don't create just temporary setbacks—they can have long-lasting economic consequences and reduce customer satisfaction worldwide.

The Project's Goals and Approach

This project aims to tackle these challenges head-on by leveraging the power of data analytics and machine learning (ML) to understand, predict, and mitigate inefficiencies in transportation systems. Our approach focuses on analyzing both traditional structured data (numerical data like flight schedules and cargo volumes) and unstructured data (visual data like aerial images of planes and ships). By integrating these two types of data, we aim to create a comprehensive view of the challenges facing transportation networks and propose actionable solutions.

Our work focuses on three primary areas:

Flight Analysis:

Using data such as flight times, weather conditions, and schedules, we employ machine learning models to predict arrival delays.

These predictions help airlines adjust their operations, optimize flight schedules, and allocate resources more efficiently.

For instance, by anticipating delays caused by adverse weather conditions, airlines can proactively reschedule flights, communicate with passengers in advance, and reduce the overall impact of disruptions.

Improved delay predictions not only enhance the passenger experience but also reduce operational costs associated with delays, such as crew overtime and fuel consumption.

Port Analysis:

Seaports handle enormous quantities of cargo daily, making them vital nodes in global supply chains. Even minor inefficiencies can accumulate, leading to significant delays and increased costs.

By analyzing data on cargo volumes, vessel types, and port call schedules, we identify congestion hotspots and inefficiencies in port operations.

For example, if data shows that bulk carriers tend to create bottlenecks during certain times of the year, port authorities can allocate additional resources, such as cranes and dock workers, to manage these ships more efficiently.

These insights can help ports improve throughput, reduce waiting times, and enhance the reliability of maritime logistics.

Image-Based Insights:

Traditional data analysis methods are limited in capturing real-time operational challenges. Aerial images provide a unique, visual perspective that complements structured data.

By applying advanced deep learning techniques like Convolutional Neural Networks (CNNs), we can detect planes on runways, identify ships in crowded ports, and monitor congestion dynamically.

For example, analyzing aerial images of an airport can reveal patterns of runway congestion that may not be evident from flight schedules alone.

Similarly, satellite images of ports can help identify areas where ships are clustered, indicating potential bottlenecks or delays.

These real-time insights allow decision-makers to respond quickly to emerging issues, improving operational efficiency.

Why This Approach Matters

What makes this approach unique is its ability to integrate both structured and unstructured data, providing a holistic understanding of transportation systems. Traditionally, studies have focused on either numerical data or image data, but rarely both. By combining these two types of data, we can offer richer insights and more practical solutions.

Examples of Integration:

For Airlines: Pairing flight delay data with real-time runway images can help airlines better manage takeoffs, landings, and gate assignments, reducing overall delays.

For Ports: Combining vessel schedules with real-time images of port activity can help port authorities dynamically allocate resources, ensuring smoother operations even during peak periods.

This integrated approach doesn’t just identify problems—it provides actionable solutions that airlines, port authorities, and logistics managers can implement to improve their operations.

Benefits and Impact

By combining data-driven predictions with actionable strategies, this project aims to make transportation systems:

More Reliable: Reduce delays and congestion through accurate predictions and real-time monitoring.

More Efficient: Optimize resource allocation, reduce operational costs, and improve throughput in airports and ports.

Better Prepared for Growth: Support the increasing demands of global trade and travel by improving the resilience and scalability of transportation networks.

Customer-Centric: Enhance passenger experiences and business satisfaction by minimizing delays and ensuring timely deliveries.

In an era where global trade and travel are growing rapidly, addressing inefficiencies in transportation is more critical than ever. This project represents a step toward a future where transportation systems are not only functional but also agile, responsive, and capable of meeting the challenges of a dynamic global economy.

Literature Review

Transportation analytics has seen remarkable progress over the years, driven by advancements in machine learning (ML) and deep learning (DL). These technologies are reshaping how we address challenges in aviation and maritime systems, offering solutions to problems like predicting flight delays, monitoring port congestion, and optimizing resource allocation. While traditional models have provided a solid foundation, recent innovations in ML and DL are enabling us to tackle increasingly complex problems with greater precision and efficiency.

Traditional Approaches in Transportation Analytics

For decades, transportation analysts relied on traditional ML models such as Decision Trees, Random Forests, and Regression Models to analyze structured data. These models have proven effective in identifying relationships between various factors like weather conditions, flight distances, and delays.

Applications in Aviation

1. **Predicting Delays**: Traditional models can identify how factors such as bad weather, airport traffic, and flight distances contribute to delays. For instance, they can determine that a snowstorm in one region is likely to delay flights not just locally but across connected routes.

2. **Operational Efficiency**: These models help airlines understand patterns in delay occurrences, enabling better planning for peak seasons and adverse weather conditions.

Applications in Maritime Operations

1. **Cargo Volume Analysis**: In ports, traditional ML models analyze cargo volumes, vessel types, and port call schedules. This helps port authorities plan for seasonal peaks, such as increased shipping activity during the holiday season.

2. **Resource Optimization**: By understanding patterns in vessel movements and port activities, ports can allocate docks, cranes, and labor more efficiently.

However, these traditional models have their limitations. They struggle to manage data with temporal dependencies — situations where the timing and sequence of events matter. For example, a delayed flight can create a domino effect, disrupting subsequent flights throughout the day. This complexity requires more advanced techniques capable of handling sequential data.

Advanced ML Techniques for Time-Based Patterns

To address the limitations of traditional models, researchers have turned to Recurrent Neural Networks (RNNs) and their more advanced variant, Long Short-Term Memory (LSTM) networks.

These models are designed to capture dependencies over time, making them particularly suited for transportation analytics where sequential patterns are critical.

Applications in Aviation

1. **Cascading Flight Delays**: RNNs and LSTMs can model how a delay in one flight can impact subsequent flights. For example, if a flight arriving late affects a connecting flight, LSTMs can predict the likelihood and extent of that delay, helping airlines adjust schedules proactively

2. **Seasonal Delay Trends**: These models can also identify patterns in delays across different seasons. For example, delays are often more frequent in winter due to adverse weather conditions, and LSTMs can help forecast these trends based on historical data.

Applications in Maritime Operations

1. **Predicting Port Congestion**: In ports, RNNs and LSTMs can analyze historical trade data to predict periods of high congestion. For instance, during peak trade seasons, these models can forecast when cargo volumes are likely to surge, allowing port authorities to plan for increased demand.

2. **Sequential Patterns in Vessel Movements**: These models can track the sequence of vessel arrivals and departures, identifying patterns that might lead to bottlenecks.

While RNNs and LSTMs are powerful for sequential data, they primarily focus on structured data such as schedules and cargo volumes. This approach leaves out the vast potential of unstructured data, particularly visual information, which can offer real-time insights into transportation systems.

Deep Learning for Image-Based Transportation Analytics

The advent of Convolutional Neural Networks (CNNs) has transformed how we work with image data. CNNs are specifically designed to detect patterns in visual data, making them ideal for tasks like object detection, classification, and scene analysis in transportation contexts.

Applications in Aviation

1. **Runway Monitoring**: CNNs can analyze aerial images of airports to detect planes on runways and taxiways. This helps identify congestion and potential bottlenecks in real-time, allowing for better gate assignments and takeoff scheduling.

2. **Airplane Detection**: Models like ResNet and EfficientNet are effective for classifying airplanes in datasets like PlanesNet. These models can handle complex images with overlapping objects, varied lighting, and small object sizes.

Applications in Maritime Operations

1. **Port Congestion Detection**: CNNs can process satellite images of ports to detect and classify ships, identify congestion, and estimate cargo volumes.

2. **Real-Time Monitoring**: Using models like YOLO (You Only Look Once), ports can detect ship locations and densities in real-time, providing immediate insights into port activity.

Popular CNN Models

ResNet: Known for its deep architecture, ResNet is effective for classifying objects in complex images, such as detecting airplanes in aerial views of airports.

EfficientNet: Balances accuracy and computational efficiency, making it suitable for large-scale image analysis tasks where speed is essential.

YOLO: Provides real-time object detection, making it ideal for dynamic environments like busy ports or airports.

These CNN-based approaches add a new layer of understanding that structured data alone cannot provide. For example, while structured data might show that a port is busy, an aerial image can reveal the exact locations of bottlenecks, allowing for more precise interventions.

Bridging the Gap: Integration of Structured and Unstructured Data

Despite significant progress in both structured and unstructured data analysis, there remains a gap in integrating these two approaches. Most research focuses on either structured numerical data or visual image data, but not both together. Combining these data types can unlock deeper insights and more comprehensive solutions.

Examples of Integration

1. **Port Operations**: Combining vessel type data (structured) with real-time images of port congestion (unstructured) can help optimize dock assignments and reduce wait times.

2. **Flight Delay Analysis**: Integrating flight schedules (structured) with runway images (unstructured) can provide a clearer picture of delays, allowing airlines to better manage operations.

This project aims to bridge this gap by integrating ML models for structured data with CNNs for image analysis. By doing so, it provides a unified framework for understanding transportation inefficiencies, leading to more effective decision-making and operational planning.

Data Understanding

To make informed decisions and solve real-world challenges in transportation, it's essential to have a deep understanding of the data we're working with. For this project, we relied on two types of datasets: tabular data (structured, numerical information) and image data (aerial visuals of planes and ships). Each dataset brings unique insights and challenges to the table, and together, they provide a more complete picture of transportation systems.

Tabular Data

1. Flight Dataset

- **Overview**:
  This dataset captures key details about flight operations, including Air Time, Distance, Departure Delay, Arrival Delay, and Weather. These variables allow us to analyze how flights are scheduled, where inefficiencies occur, and what factors contribute to delays.

- **Target Variable**:
  Our main focus is on ARR_DELAY—the arrival delay in minutes—which is a critical metric for understanding flight efficiency and passenger satisfaction.

- **Insights**:
  - **Seasonal_Trends**:
    Delays spike during the winter months due to bad weather like snowstorms, icy runways, and reduced visibility. These conditions make it harder for flights to stay on schedule, highlighting the importance of weather-related variables.
  - **Long-Haul-Flights**:
    Longer flights tend to face more significant delays, as seen in scatter plots comparing Air Time and Distance. This is likely because longer routes have more touchpoints and dependencies, leaving little room for recovery when things go wrong.
  - **Role_of_the_weather**:
    Unsurprisingly, weather emerged as a major factor. Extreme conditions like thunderstorms or heavy fog often lead to delays, making weather forecasting a critical part of flight management.

2. Port Dataset

- **Overview**:
  This dataset provides insights into port operations, focusing on variables such as Import_Cargo (cargo volume), Vessel_Type (e.g., tankers, bulk carriers), and Port_Calls (number of ships arriving). These metrics give us a clear view of how busy ports are and where bottlenecks might occur.
- **Insights**:
  - **The Role of Bulk Carriers and Tankers**:
    Bulk carriers and tankers are often responsible for significant congestion because they require more time and space to unload compared to smaller vessels. They're a key factor in how efficiently a port operates.
  - **Seasonal Peaks**:
    Cargo volumes tend to hit their highest levels in Q3 (July–September), which aligns with global trade cycles and holiday season preparations. This trend highlights the need for ports to plan resources carefully during these busy months.
- **Visualizations**:
  - **Monthly Trends**:
    Line charts of monthly cargo volumes show a clear seasonal pattern, helping us anticipate peak activity periods.
  - **Correlation Insights**:
    Correlation matrices reveal how vessel types influence congestion, making it easier to pinpoint which ships contribute the most to delays.

Image Datasets

1. PlanesNet

- **Description**:
  PlanesNet is a dataset of over 32,000 aerial images, each labeled to show whether or not a plane is present. These images provide a bird's-eye view of airport operations and are particularly useful for understanding runway congestion and activity levels.
- **Challenges**:
  - **Small Objects**:
    In high-resolution aerial images, airplanes appear relatively small, making them harder to detect accurately.

  - **Lighting Conditions**:
    Variations in lighting (e.g., daytime vs. nighttime or cloudy vs. sunny skies) add complexity to the detection process.
  - **Overlapping Planes**:
    Multiple planes parked close together or on a runway can make it difficult for models to differentiate between them.
- **Applications**:
  This dataset is especially helpful for monitoring airport congestion in real time. For instance, if too many planes are queued on a runway, it signals potential delays. Insights like these can help optimize gate assignments and improve turnaround times.

2. Ships in Aerial Images

- **Description**

  This dataset comprises annotated aerial images with bounding boxes around ships, making it well-suited for tasks such as object detection and ship classification. The images encompass a variety of scenarios, ranging from sparsely populated docks to highly congested ports, providing a comprehensive view of port activity.
- **Insights**:
  - **Real-Time Congestion Monitoring**:
    Analyzing these images enables the identification of areas with high ship density, indicating potential congestion. This insight is crucial for optimizing berth assignments, improving traffic flow, and minimizing delays in cargo handling.
  - **Ship Classification**:
    The dataset facilitates the classification of various ship types, including tankers, cargo vessels, and passenger ships. Understanding the types and numbers of ships present in a port allows for better prioritization of movements and more efficient allocation of resources.
- **Applications**:
  Insights from this dataset are key for managing port traffic. For example, operators can use the data to decide which ships should dock first or how to best distribute resources to handle peak traffic efficiently.

The tabular data provides a detailed look at the numbers behind flight and port operations, highlighting trends, patterns, and areas for improvement. On the other hand, the image datasets offer a visual perspective, giving us real-time insights into congestion and activity levels. Together, these datasets enable a well-rounded analysis of transportation systems, paving the way for smarter and more efficient decision-making.

Methodology

To address the challenges in transportation systems, we followed a structured methodology that involved data preprocessing, model development, and evaluation. Our approach integrated both structured tabular data and unstructured image data to provide a comprehensive view of inefficiencies in flights and ports. Each step was carefully designed to ensure the models could deliver accurate and actionable insights.

Data Preprocessing

Before diving into modeling, the data underwent extensive cleaning and preparation. This process was crucial to ensure the reliability and accuracy of the models. The preprocessing steps varied for the two types of data we worked with: tabular data and image data.

Tabular Data

Tabular datasets provided numerical insights into flight delays, cargo volumes, and port activities. However, they required significant preprocessing to ensure their quality and suitability for machine learning models.

Handling Missing Values

Gaps in data, especially in critical fields like weather and flight schedules, posed a challenge.

- **Weather Data**: Missing temperature or precipitation values were filled using interpolation techniques. For instance, if a specific time point was missing, we estimated it using the average of neighboring values to maintain data continuity.
- **Flight Schedules**: Missing entries for delays were treated based on historical averages or median values to avoid introducing bias.

Feature Engineering

To capture seasonal patterns and cyclical trends, we created new features such as:

- **Quarter of the Year (e.g., Q1, Q2)**: Capturing seasonal variations like winter delays or summer peaks.
- **Day of the Week**: Understanding whether delays were more frequent on weekends or weekdays.
- **Holiday Flags**: Marking flights or shipments occurring on public holidays to see if demand surges impacted delays.
- **Interaction Features**: We also created interaction terms, such as combining weather conditions with flight distance, to capture complex relationships.

Transforming Skewed Data

Certain variables, like Cargo Volume, had extreme values that could disproportionately influence the model's predictions.

We applied logarithmic scaling to normalize these variables, reducing the impact of outliers and making patterns easier for the model to learn.

Standardization and Scaling

Numerical variables like Distance and Flight Time were standardized to ensure consistent ranges, which is essential for machine learning algorithms sensitive to feature scales (e.g., Gradient Boosted Trees).

Categorical Encoding

Non-numerical fields like Airport Codes and Vessel Types were converted into numerical representations using one-hot encoding or target encoding, ensuring they could be used effectively in the models.

Image Data

Image data required a completely different set of preprocessing techniques tailored to the needs of deep learning models, particularly Convolutional Neural Networks (CNNs).

Resizing

All images were resized to 224x224 pixels to ensure compatibility with models like ResNet and YOLO. This size was chosen to balance computational efficiency with the need to retain enough detail for accurate object detection and classification.

Normalization

Pixel values were scaled to a range of 0 to 1 by dividing by 255, making the data easier for the models to process. This step ensures that models converge faster during training.

Data Augmentation

To improve model generalization and robustness, we applied the following augmentation techniques:

- **Flipping**: Simulated different viewing angles, allowing the model to recognize objects regardless of their orientation.
- **Rotation**: Helped the model understand variations in object placement, such as ships docked at different angles.
- **Brightness Adjustments**: Prepared the model for varying lighting conditions, such as bright sunlight or overcast skies.

- **Cropping**: Simulated partial views of objects, ensuring the model could still identify airplanes or ships even if parts were obscured.
- **Zooming**: Allowed the model to handle variations in object size within images.

Bounding Box Validation

For datasets like Ships in Aerial Images, bounding box annotations were manually verified for accuracy. This ensured that the models trained on correctly labeled data.

Image Class Balancing

If certain classes (e.g., airplane vs. no airplane) were underrepresented, techniques like oversampling or synthetic data generation (e.g., GANs) were used to ensure balanced class distributions.

Model Development

After preprocessing the data, we developed models tailored to the specific challenges of flight and port inefficiencies.

1. **Tabular Data Models**

- **Gradient Boosted Trees and Random Forests:**

  Used for initial predictions of flight delays and port activity patterns.

  These models provided baseline results and identified key feature relationships.

- **RNN with LSTM:**

  Designed to capture sequential dependencies in flight delays, such as cascading effects caused by earlier disruptions.

  Time-series data, such as flight schedules, was transformed into sequences of events, allowing the model to predict how delays would propagate.

2. Image-Based Models

- **ResNet:**

  Fine-tuned for binary classification tasks, such as detecting the presence of airplanes in aerial images.

  This model was chosen for its ability to handle complex imagery while maintaining high accuracy.

- **YOLO (You Only Look Once):**

  Used for real-time object detection in the Ships in Aerial Images dataset.

  YOLO provided precise bounding boxes around ships, enabling real-time monitoring of port congestion.

Model Evaluation

We evaluated the models using a combination of metrics to ensure their reliability and applicability in real-world scenarios.

Tabular Data Models:

- **Metrics**: R^2, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

  LSTMs were evaluated using time-series-specific metrics, such as Mean Absolute Error (MAE) for sequential predictions.

Image-Based Models:

- **Metrics**: Accuracy, Precision, Recall, and F1-Score.

  Object detection models like YOLO were evaluated based on Intersection over Union (IoU) scores to measure the accuracy of bounding boxes.

Data Modeling

In order to address the challenges faced by transportation systems, we developed models specifically tailored to each type of data—structured tabular data and unstructured image data. Additionally, we implemented hybrid approaches that combined insights from both types of data to provide a more comprehensive understanding of inefficiencies. Our approach to data modeling involved careful selection of algorithms, iterative testing, and continuous refinement to achieve accurate and reliable predictions.

Tabular Data Models

For structured data, which included flight schedules, weather conditions, cargo volumes, and vessel movements, we focused on predicting flight delays and port activity. These models were designed to capture patterns and relationships within the numerical data.

- **Random Forests and Gradient Boosted Trees**

  Random Forests:

  These ensemble models are robust and effective at handling large datasets with multiple features. By combining the results of many decision trees, Random Forests reduce the risk of overfitting and provide reliable predictions.

  Feature Importance:

  One of the key advantages of Random Forests is their ability to rank the importance of different features. For example, the model highlighted that weather conditions, flight distance, and departure time were the most critical factors influencing flight delays.

  Interpretability:

  The insights gained from feature importance helped us understand which factors needed more attention in operational planning.

  Port Analysis:

  In port activity predictions, Random Forests identified the impact of vessel type, cargo volume, and seasonal trends on port congestion.

  Gradient Boosted Trees:

  These models improve prediction accuracy by correcting errors in sequential steps. Each new tree focuses on the mistakes made by the previous ones, leading to more refined predictions.

Advantages:

Gradient Boosted Trees are highly effective for datasets where subtle patterns need to be captured.

Applications:

In flight delay prediction, Gradient Boosted Trees helped fine-tune the results, providing more accurate forecasts of arrival delays based on combinations of factors like weather severity and flight distance.

- **LSTM Networks (Long Short-Term Memory)**

Handling Sequential Data:

Flight schedules are inherently sequential, meaning a delay in one flight can affect subsequent flights. To capture these cascading effects, we used LSTM networks, a type of Recurrent Neural Network (RNN) designed to handle time-dependent data.

Why LSTM:

Unlike traditional models, LSTMs can maintain long-term dependencies, making them ideal for predicting how delays propagate through a flight schedule.

Application:

The LSTM model was trained to predict how an initial delay might impact future flights within an airline's network. For example, a delay in a morning flight might lead to a ripple effect of delays throughout the day.

Sequential Insights:

LSTM networks captured patterns such as:

Peak Delay Times:

Identifying periods of the day or week when cascading delays are more likely.

Crew and Aircraft Availability:

Understanding how delays affect crew assignments and aircraft turnaround times.

Visualization of Predictions:

Time-series graphs were generated to show how the LSTM model predicted delays over a sequence of flights, providing clear insights into potential bottlenecks.

Image-Based Models

For unstructured image data, we used advanced deep learning techniques to analyze aerial images of airports and ports. These models provided visual insights that complemented the numerical data, enabling real-time monitoring and analysis of transportation operations.

- **ResNet (Residual Network)**

  Architecture:

  ResNet is a powerful Convolutional Neural Network (CNN) known for its ability to train very deep networks without suffering from degradation (vanishing gradients).

  Fine-Tuning:

  We fine-tuned ResNet on the PlanesNet dataset to classify images as containing a plane or not. This approach allowed the model to handle challenges like:

  Small Object Size:

  Airplanes that occupy only a small portion of the image.

  Overlapping Objects:

  Planes on busy runways with multiple objects in close proximity.

  Varied Conditions:

  Different lighting and weather conditions affecting image quality.

  Applications:

  1. Runway Congestion Detection: Identifying planes on runways and taxiways to monitor congestion in real time.
  2. Operational Efficiency: Providing airport authorities with real-time insights into ground operations, allowing for better gate assignments and takeoff scheduling.

  YOLO (You Only Look Once)

  Real-Time Object Detection:

  YOLO is known for its speed and accuracy in detecting multiple objects within a single image.

  Port Congestion Analysis:

  Applied to the Ships in Aerial Images dataset, YOLO detected and classified ships with high precision, even in crowded or cluttered environments.

Bounding Boxes: The model generated bounding boxes around detected ships, enabling port authorities to visualize congestion hotspots.

Applications:

1. Dynamic Resource Allocation: Ports could use these insights to dynamically allocate resources, such as assigning more cranes to areas with high ship density.
2. Monitoring Vessel Types: Classifying different types of vessels (e.g., bulk carriers, tankers) to better understand traffic patterns.

Hybrid Models

To maximize the insights gained from both tabular and image data, we developed hybrid models that combined features from both sources. This approach provided a more comprehensive analysis of transportation inefficiencies.

Integration of Features

Port Analysis Example:

Combined tabular data (e.g., vessel type, port call schedules) with image-derived features (e.g., ship density and location).

Benefits:

Captured both historical trends (from tabular data) and real-time conditions (from image data).

Improved predictions of congestion by incorporating up-to-date visual information.

Enhanced Predictions and Decision-Making

Airline Operations:

Merged flight schedules with runway imagery to identify potential bottlenecks and predict delays more accurately.

This combination helped airlines manage gate assignments and optimize flight turnaround times.

Port Operations:

Hybrid models provided real-time insights into port congestion, enabling proactive decision-making and resource allocation.

For instance, if the image data showed a sudden increase in ship density, port managers could reassign docking slots to alleviate congestion.

Evaluation Metrics

To assess how well our models performed, we used metrics suited to the specific tasks:

- **Classification Metrics** (for image-based tasks):
  - **Accuracy**: How often the model got the predictions right.
  - **Precision**: How reliable the positive predictions were (e.g., correctly identifying planes or ships without false alarms).
  - **Recall**: How many actual positives were correctly identified (e.g., ensuring all planes or congested areas were detected).
  - **F1-Score**: A balanced measure that combines precision and recall, especially useful when false positives and false negatives carry similar costs.
- **Regression Metrics** (for tabular data tasks):
  - **R^2 (Coefficient of Determination)**: How well the model explained the variability in the target variable, like delays or cargo volumes.
  - **Mean Squared Error (MSE)**: The average squared difference between predicted and actual values, penalizing larger errors.
  - **Root Mean Squared Error (RMSE)**: Like MSE, but easier to interpret since it's in the same units as the target variable.

This methodology brought together the best of both worlds—structured data for understanding trends and patterns, and image data for real-time insights. By carefully preparing the data and using specialized models for each task, we were able to tackle challenges like flight delays and port congestion effectively. The hybrid approach further amplified the value of the models, showing how combining different types of data can lead to richer and more actionable insights.

Results

The results of this project demonstrate the potential of combining data-driven approaches with advanced technology to tackle inefficiencies in transportation. By analyzing flight delays, port activities, and congestion through both tabular and image data, we achieved meaningful insights that could directly benefit operational decision-making. Here's a breakdown of the key findings:

1. Flight Analysis

For flight delays, the goal was to predict arrival times accurately and understand the main factors contributing to delays.

- **Model Performance**:
  Our models achieved an R^2 of 0.85, meaning they explained 85% of the variation in arrival delays. This high accuracy shows that the models successfully captured the critical elements influencing flight delays.
- **Key Insights**:
  The two most significant factors driving delays were weather and flight distance. Extreme weather conditions, like heavy rain or snow, were a major culprit, especially during winter months. Similarly, longer flights tended to have higher delays, likely because they depend on more variables, such as air traffic conditions and scheduling constraints.
- **Seasonal Trends**:
  Winter emerged as a particularly challenging time for airlines, with delays spiking during this season. This insight aligns with real-world challenges, such as icy runways and increased passenger volumes during the holiday season. These findings can help airlines prepare better, such as scheduling additional buffer times for winter flights.

2. Port Analysis

The port analysis focused on predicting cargo volumes, understanding congestion patterns, and improving resource management.

- **Model Performance**:
  Our regression models explained 80% of the variance in cargo volumes, providing a strong understanding of port activity. The hybrid models, which combined tabular data and image features, performed even better, offering a deeper view of congestion and operational bottlenecks.
- **Hybrid Model Benefits**:
  By integrating data from both sources, the hybrid models brought together the best of both worlds. For example, while the tabular data captured historical trends, the images

provided real-time information about ship density and congestion levels. Together, they created a fuller picture of what was happening at the port.

- **Practical Takeaways**:
  Cargo volumes peaked in Q3, coinciding with global trade cycles and holiday demand. Ports can use this information to allocate resources more effectively during busy periods. Additionally, the insights highlighted how certain vessel types, such as bulk carriers and tankers, disproportionately contribute to congestion, offering clear targets for optimization.

3. Image-Based Analysis

Using deep learning models, we analyzed aerial images to gain visual insights into congestion at airports and ports. These results brought real-time context to the structured data.

PlanesNet Analysis

- **Performance**:
  Our ResNet model achieved an impressive 92% accuracy in detecting planes from aerial images. This level of accuracy highlights the model's ability to handle challenges like small object sizes and overlapping planes.
- **Interpretability**:
  To ensure the model's predictions were transparent, we used Grad-CAM (Gradient-weighted Class Activation Mapping) to generate heatmaps. These heatmaps showed the exact areas of the image the model focused on, such as the body of the plane or its shadow, making the decision-making process easy to understand and trust.
- **Applications**:
  This analysis is invaluable for monitoring airport congestion. For example, detecting clusters of planes on runways or parking areas can help identify potential bottlenecks, enabling airport staff to act quickly and avoid delays.

Ships in Aerial Images Analysis

Performance

The CNN model with ResNet architecture, trained using the SGD optimizer, demonstrated exceptional performance, making it the best-performing model with an accuracy of 92%, an F1-score of 96%, and an AUC of 0.55. These metrics highlight its robust ability to accurately identify and classify ships within complex aerial images, even in densely populated port environments.

Real-Time Insights

The ResNet-powered model's advanced feature extraction and real-time detection capabilities enabled effective monitoring of dynamic port activity. For instance, it could swiftly identify areas with high ship density, providing valuable insights to anticipate and address potential congestion.

Applications in Port Management:
These insights are incredibly useful for port authorities. Knowing exactly where ships are and how crowded specific areas are can help operators prioritize berthing schedules, reduce wait times, and improve overall efficiency.

This project delivered meaningful results that highlight the potential of advanced analytics in solving real-world transportation problems:

- **For Flights**: The models identified weather and distance as key predictors of delays, with seasonal trends providing actionable insights for better scheduling and resource allocation.
- **For Ports**: Hybrid models combining tabular data and image-based insights proved highly effective, offering both historical context and real-time congestion data.
- **For Aerial Images**: Deep learning models like ResNet and YOLO demonstrated their ability to analyze visual data with impressive accuracy, enabling real-time monitoring of airport and port operations.

These results underscore the importance of combining structured and unstructured data to gain a complete understanding of transportation systems. The findings not only help explain inefficiencies but also provide actionable solutions, from optimizing flight schedules to improving port resource management.

Discussion

This study shows how combining different types of data—structured numbers like flight times and port activities, and unstructured visuals like aerial images—can help us tackle inefficiencies in transportation systems. By blending machine learning (ML) and deep learning (DL), we gained a deeper understanding of what causes delays and congestion and how to fix them. ML models did a great job working with tabular data, while CNNs excelled at analyzing images, and together, they gave us a complete picture of the issues.

Making a Difference for Airlines

For airlines, the insights from this study can directly improve operations. Our ML models identified weather and distance as the main culprits behind flight delays. This isn't surprising—bad weather, especially during winter, and longer flight routes often create challenges. But what's exciting is that these models can help airlines prepare better. For example:

- Airlines can schedule more buffer time for long-haul flights or during winter months when delays are more common.
- Predictive models can give passengers more accurate delay estimates, improving their travel experience.
- Real-time congestion insights, like detecting overcrowded runways through CNNs, can help airlines and airports adjust operations on the fly—literally.

This means fewer cascading delays, smoother schedules, and happier passengers.

Improving Port Operations

Ports are another area where these insights can make a big impact. Our models showed that cargo volumes peak in Q3, aligning with the holiday season and global trade cycles. Bulk carriers and tankers were also flagged as significant contributors to congestion because they take up more space and time. By understanding these patterns, port managers can:

- Plan better for busy seasons, ensuring they have enough staff and equipment.
- Prioritize ships carrying time-sensitive cargo for quicker unloading.
- Optimize berthing schedules to reduce wait times for larger vessels.

The hybrid models, which combined numerical data with aerial imagery, were particularly powerful. They not only revealed historical trends but also gave a live view of what was happening in the port, like identifying which zones were becoming overcrowded. This kind of dynamic insight allows managers to act quickly, redirect ships, and keep operations running smoothly.

The Bigger Picture

This study highlights just how powerful it is to combine structured and unstructured data. By blending historical trends with real-time insights, we can do more than just understand what's happening—we can take proactive steps to prevent problems before they escalate. For example:

- Airlines can better manage schedules and reduce the ripple effect of delays.
- Ports can allocate resources more effectively during peak times and reduce bottlenecks.

But the potential goes beyond just flights and ports. This approach could be applied to other areas, like rail networks, urban traffic management, or even logistics hubs like warehouses. Anywhere there's a mix of data types and operational complexity, this methodology could provide similar benefits.

Challenges Along the Way

Of course, like any project, this study wasn't without its challenges:

1. **Data Gaps**: Missing values in flight and port data required careful handling, and some aerial images varied in quality, making it harder for models to interpret them.
2. **Computational Power**: Deep learning models, particularly CNNs like ResNet, required significant computing resources to train. This could be a hurdle for organizations without access to high-performance computing.
3. **Real-Time Application**: While the models worked well in a research setting, deploying them in real-time environments (like live monitoring of airports or ports) would require further integration and testing.

Future Scope

This study is just the beginning. With advancements in technology—like better satellite imagery and IoT sensors—we can collect even richer data, making these models even more accurate and effective. Future possibilities include:

- Real-time systems that combine predictive models with live data feeds to dynamically adjust operations.
- Extending this approach to other transportation networks, like trains or urban traffic, to make systems smarter and more efficient.
- Collaborating with industry stakeholders to implement these solutions on a larger scale.

This study showed the value of bringing together different types of data and technologies to solve real-world problems in transportation. By combining ML for tabular data and DL for images, we've created a way to not only understand inefficiencies but also propose actionable solutions. From helping airlines reduce delays to giving ports the tools to manage congestion, these insights have the potential to transform how we manage transportation systems—making them more efficient, reliable, and ready for the future.

Conclusion

This project showed how powerful it can be to combine machine learning (ML) and deep learning (DL) techniques to tackle real-world challenges in transportation. By using both structured data, like flight schedules and cargo volumes, and unstructured data, like aerial images, we were able to uncover patterns and inefficiencies that wouldn't have been obvious otherwise. This integration gave us a much richer and clearer understanding of the problems affecting aviation and maritime operations.

In aviation, our models were able to predict flight delays with impressive accuracy, highlighting key factors such as weather, flight distances, and seasonal trends. This kind of insight is invaluable for airlines, as it can help them improve scheduling, reduce cascading delays, and create better experiences for passengers. Using advanced models like LSTMs, we even captured how delays in one part of the system ripple out to affect others—something traditional methods often struggle to do.

On the maritime side, analyzing port data revealed clear trends, like how certain vessel types and seasons contribute to congestion. Combining this with image data added a whole new layer of understanding. For example, our models could visually identify and monitor congestion in ports, helping port authorities allocate resources more effectively and keep things running smoothly, even during peak seasons.

One of the standout aspects of this project was using CNNs to analyze aerial images. With tools like ResNet, we were able to detect airplanes and ships in real time, giving insights into runway and port activity that could help with congestion management. Even more importantly, visual tools like Grad-CAM allowed us to explain how these models were making their decisions, ensuring the results weren't just accurate but also easy to trust and understand.

This project was the seamless integration of structured and unstructured data. This hybrid approach enabled us to combine the strengths of both data types:

- **Structured Data**: Provided historical patterns, statistical relationships, and quantifiable trends in flight delays and port congestion.
- **Unstructured Data**: Offered real-time visual insights, capturing current conditions and bottlenecks that structured data alone could not reveal.

By merging these insights, we developed models that delivered a comprehensive view of transportation challenges. This integration allowed us to generate more accurate predictions, actionable strategies, and practical solutions that stakeholders could implement in real-world scenarios.

In the end, this project highlights how technology like ML and DL can transform the way we approach transportation issues. With the right data and tools, we can reduce delays, optimize resources, and make global trade networks more reliable and efficient. As more data becomes available and technology continues to improve, the potential for these methods will only grow, making them essential for the future of transportation.

References

1) Chollet, F. (2018). Deep Learning with Python | "A foundational book on deep learning concepts and techniques, including LSTM networks for sequential data and CNNs for image analysis.". https://www.manning.com/books/deep-learning-with-python

2) "Planes in Satellite Imagery | Kaggle." https://www.kaggle.com/datasets/rhammell/planesnet.

3) Kaggle. (2023). Ships in Aerial Images Dataset. | "An annotated dataset for ship detection and classification tasks, essential for real-time port congestion analysis. https://www.kaggle.com/datasets/siddharthkumarsah/ships-in-aerial-images

4) Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. | "The foundational paper introducing LSTM networks, critical for modeling sequential data such as flight delays." https://doi.org/10.1162/neco.1997.9.8.1735

5) YOLO Object Detection. (2020). You Only Look Once (YOLO) Algorithm. "A widely used algorithm for real-time object detection, applied to ship and airplane detection in this project." https://pjreddie.com/darknet/yolo/