Predictive analysis to associate damage and tolls to a driver/lane/trailer

Ananth Mohan, Farha Shireen, Hsueh-Ning Chao, Shih Min Lin, Tanvir Ahmed Farook,
Yao Liu, Yijun Wang, Dr. Shoaib Khan
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
mohan52@purdue.edu , lnuf@purdue.edu, chao77@purdue.edu, lin1944@purdue.edu,
tfarook@purdue.edu, liu4091@purdue.edu, wang6665@purdue.edu, khan180@purdue.edu

ABSTRACT [2.5 points]

Predictive analytics has the potential to transform cost management and damage tracking in the logistics industry by integrating telematics, toll records, and automated damage assessment systems. This study focuses on designing a predictive system to streamline cost attribution, improve operational transparency, and reduce billing disputes for a trailer-as-a-service provider. The importance of this work lies in its ability to address inefficiencies caused by manual processes, data gaps, and a lack of real-time insights. Using machine learning models and AI-powered image recognition, we developed a framework that assigns costs to specific drivers, lanes, or trailers while automating damage detection. The findings indicate that the integration of predictive tools not only enhances operational efficiency but also builds customer trust through accurate billing and transparent processes.

Keywords: predictive modeling, computer vision, telematics, cost attribution, logistics, AI damage detection, billing transparency

INTRODUCTION [20 points]

The global third-party logistics (3PL) market size is anticipated to grow from USD 1082.45 billion to USD 2230.96 billion in 10 years. The market will experience rapid growth due to technological advancements in third-party logistics (3PL) during the forecast period (Yahoo Finance, 2024). One of the service third party logistics provide is transportation service, which the company own their trailer and distribute goods for the customer. For asset-based 3PL, fleet management could be very complicated. There are 3 types of trailer management -- traditional leasing, lease-to-own or direct ownership. However, direct owning trailer assets demand huge initial investment and maintenance cost, which makes capacity very non-flexible, alongside with potential loss on asset downtime. Some 3PL opt for traditional leasing to avoid the maintenance cost, but it failed to meet the flexibility on peak and low-capacity periods. As a result, traditional 3PL logistics business model face challenges on balancing capacity and efficiency.

Trailer as a Service (TaaS) provides a flexible solution on this challenge. Companies subscribe for capacity on the trailer, and TaaS takes care of maintenance. This solves the pain point on non-flexible, seasonal capacity and huge capital investment, and expand the opportunities to not only 3PL but brokers and carriers. While providing solutions on capacity, inaccurate billing and untracked operational cost of tolls and damages expenses become challenges for our client running the TaaS service. As the subscribers could utilize the trailer by assigning to different carriers, tracking tolls and damage expenses become challenging, given our client only has subscribers and asset list rather than the actual parties operate the

trailers. This could lead to inaccurate billing and could further strained relationships with customers and carriers.

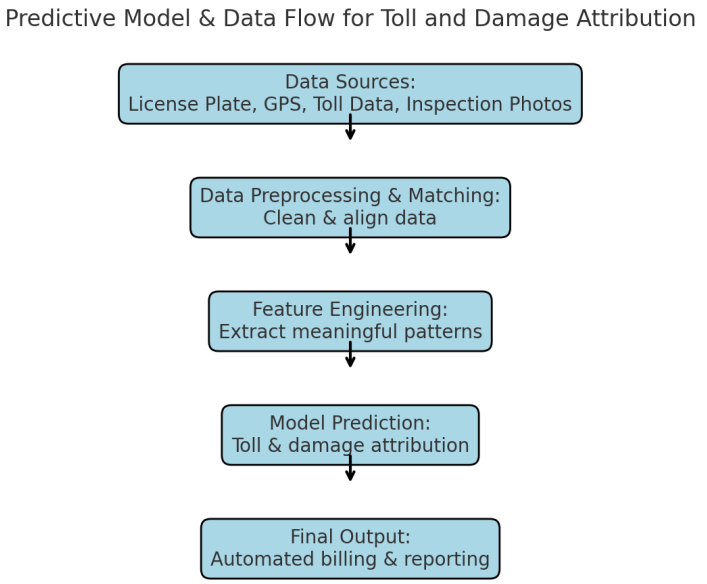


Figure1. Predictive Model & Data Flow

In our study, the motivating business problem this paper focuses on is to build efficient workflow and model help attributing the tolls and damage to the appropriate party in a timely manner without relying on data provided from service subscribers.

Our research will focus on:

1. What approaches from predictive modeling along with data analytics systems should be adopted to enhance both accuracy and efficiency of toll attribution operations within TaaS systems?

2. Establish methods that enhance trailer damage monitoring together with attribution processes despite existing inspection irregularities and trailer exchange frequency.

The analysis will build predictive tools to automate toll expense allocation and damage inspection functions that support operational requirements of the business. Our research will review different techniques to pair toll payments with proper customers despite restricted data availability through the combination of license plate recordings timestamp records and GPS data.

An algorithm needs to be developed to process large datasets effectively throughout making correct predictions with limited available information. Our research investigates the application of computer vision technology for pre-trip and post-trip inspection photo comparison automation which helps detect possible manual inspection misses of vehicle damages. With integrating data from tolls agency and GPS data, our client will be able to automate the tolls attribution and bill the damage expense on the appropriate parties, which further reduced unnecessary overhead.

The following paper structure includes: Section 2 contains a literature review about transportation industry predictive models alongside damage detection techniques and toll attribution methods. Our

proposed methodology structure comprises data preprocessing alongside feature engineering and model selection which appears in Section 3. Our experimental results with model descriptions appear in Section 4 following our method implementation description. The evaluation of our models' performance and practical applications and possible enhancement opportunities takes place in Section 5. Section 6 delivers the paper's summary including our study findings along with their effects on TaaS business and prospective research paths. The research we conduct addresses crucial operational challenges of the TaaS model with the goal to boost digital transformation within the transportation and logistics industry through efficient and transparent operations.

LITERATURE REVIEW [30 points]

Tolls Attribution

Cost allocation in logistics can be complex due to various operational factors and data limitations. In our case, matching toll data in a timely manner poses a challenge, especially when customers have limited information about carriers or brokers. Sheng Xu et al. (2020) introduced a time-driven activity-based costing (TDABC) model integrated with a shared logistics platform, designed to facilitate real-time data updates and reduce implementation costs, making cost allocation more efficient and adaptable.

Damage Attribution

The use of transfer learning on damage detection on containers has been research widely, especially for MobileNetV2 model. This has been concluded by (Zixin Wang et al., 2021) and (Pavel Cimili, et al., 2022) that MobileNetV2 has advantages in multiple class of damage detection and it is useful in large scale container inspection scenerios. The former compared transfer and semi-supervised approaches by testing two separate models (for the lower and the upper part), and it was concluded that semi-supervised training could not outperform transfer learning model due to the complex structure of the trailer surface and its defects. They suggest that future improvement is needed for the classification of multiple damage classes.

The potential of MobileNetV2 model for multiple types of damage detection is still been researched. Zixin Wang et al., 2021 proposes a multitype damage detection model for containers based on MobileNetV2. They performed on-stie experiment on deploying model on the mobile terminal obtains images through the smartphone camera for real-time damage detection. They concluded that experiment results show that the multitype container damage detection model can give the corresponding damage types and prediction results. However, it is still necessary to quantify the degree of damage to the container according to the severity of the injury to the container and support the intelligent decision-making of container damage.

This provides the groundwork for our paper as we will focusing on attributing tolls cost based on geographical data and identifying damage classification. Our goal is integrating this information and match them to the appropriate party.

We summarize our findings in Tables 1 and Tables 2.

Table 1: Key papers and identified research gaps

Study	**Insights**	**Research Gap**
(Sheng Xu et al., 2020)	They introduced a decision support platform based on shared logistic	

	platform, which can be used to collect and integrate data in the process of time-driven activity-based costing.	
(Pavel Cimili, et al., 2022)	MobileNetV2 based on transfer learning is capable of damage detection if there is enough training data.	The model has the potential to applied on not just for binary classification but also for multiple class damage detection.
(Jiahao Chen et al., *n.d.*)	This paper proposes an improvement to the YOLOv5 model based on the Transformer self-attention mechanism for container damage detection, demonstrating superior performance compared to commonly used object detection algorithms.	Assessing the severity of multiple damaged areas in containers still requires enhancement.
(Zixin Wang et al., 2021)	This paper proposes a multitype damage detection model for containers based on MobileNetV2, which has excellent advantages in largescale container inspection scenarios.	A real-time monitoring system will needed to be developed based on port IP network cameras and integrated into the port management system. Also it is necessary to quantify the severity of damage to the container.

Table 2: Relation of our study to other academic papers

Study	**Paper Aspect**					
	Tolls		Damage			
	TDABC model	Shared Logistics Platform	Mobile NetV2	Semi-Supervised Learning	MultipleType damage classification	Model Enhancement
(Sheng Xu et al., 2020)	V	V				
(Pavel Cimili, et al., 2022)			V	V		
(Jiahao Chen et al., *n.d.)*					V	V
(Zixin Wang et al., 2021)			V		V	
Our Study	V		V	V		

Other Important Links (To be summarised and link to be deleted)

- https://www.eliftech.com/insights/ai-in-logistics-explained/
- https://iopscience.iop.org/article/10.1088/1742-6596/1880/1/012012/pdf
- Leveraging AI-Driven Decision Intelligence for Complex Systems Engineering
- https://iopscience.iop.org/article/10.1088/1742-6596/1880/1/012012/meta
- https://link.springer.com/chapter/10.1007/978-981-16-5157-1_25
- https://arxiv.org/abs/2403.06674
- https://ieeexplore.ieee.org/abstract/document/9751889?casa_token=gjccDuJjNj4AAAAA:dFoEemZtNCYvz4UJRYHH1td8rqyk8G43S4FXpwJk7URLxKj1e6U91J8HFK2rgQ-7yWDMPtYAq4M

DATA [5 points]

Describe the data set you are using and where it can be found if it's public (e.g. Kaggle.com). If it is a proprietary, such as a client's data set, go ahead and fill out the data dictionary type table. If you have many features because of they are dummy variables, just create one row in your table that describes what it measures and how many factor levels there are. If you have a lot of features that are not categorical, then you can describe them in words without a table. For example, "*in this study we had weather measurements (e.g., degrees Celsius, humidity, etc.). There are competitor measures (e.g., market share, brand awareness, etc.)….*"

If you have a table it should describe the data type, provide the units of measure, and provide an informative description. Do not begin a section with a figure or table. Begin with words and then reference the table/figure.

The dataset includes three main tables: **Asset Location**, **Inspections**, and **Tolls**.

- **Asset Location**: Tracks trailer information like unique identifiers (asset_vin), geospatial data (position), current motion status (asset_motion_status), and time stamps (reported_time, created_time). It also includes provider data (telematic_provider).
- **Inspections**: Contains details about inspections with variables such as inspection ID (id), trailer ID (vin), inspection type (inspection_type), and time stamps (start_time, end_time). It also stores metadata like user IDs (created_by, updated_by) and inspection data in JSON format.
- **Tolls**: Records toll event data, including date and time (Posted_Date, Invoice_Date), toll detection method (Read_Type), vehicle ID (Device_Plate_Id), toll charges (Toll_Charge), and plaza information (Entry_Plaza, Exit_Plaza). It also tracks disputes (Dispute_Status, Dispute_Reason) and account associations.

Table 1: Asset Location

Variable	Type	Description
asset_vin	Categorical	Unique identifier for each trailer
position	Geospatial	Geospatial data refers to the trailer spot
location	Text	The corresponding address of the geospatial data
asset_motion_status	Categorical	The current status of the trailer
asset_name	Categorical	The specific identifier given to the trailer within the company; corresponding to trailer number in Inspections table, Vehicle Number in Tolls table
organization_anon	Categorical	
telematic_provider	Categorical	The providers of the geospatial and location data
reported_time	Time	The time corresponding to the trailer position, location and status
created_time	Time	

Table 2: Inspections

Variable	Type	Description
created_by	Categorical	Unique identifier for each asset
updated_by	Categorical	Geospatial data refers to the trailer spot

created_time	Time	The corresponding address of the geospatial data
updated_time	Time	The current status of the trailer
id	ID	The specific identifier given to the trailer within the company
configuration_id	ID	
data	JSON	The providers of the geospatial and location data
is_deleted	Boolean	The time corresponding to the trailer position, location and status
vin	Categorical	Unique identifier for each trailer
inspection_type	Categorical	Indicate Pre or Post repeort
organization_id	Categorical	
document_id	ID	Unique identifier for each inspection reports
unit_number	N/A	
trailer_number	Categorical	The specific identifier given to the trailer within the company ; corresponding to asset_name in Asset Location table
start_time	Time	Inspection started time
end_time	Time	Inspection ended time
inspection_number	ID	Unique identifier for each inspection reports
location_id	Categorical	

Table 3: Tolls

Variable	Type	Description
Posted_Date	Time	The date the toll was incurred
Invoice_Date	Time	The date the toll was billed
Source	N/A	
Read_Type	Categorical	The type detected by the toll pass
Transponder_Status	Categorical	
Device_Plate_Id	Categorical	License plate for each vehicle
Vehicle_Number	Categorical	The specific identifier given to the trailer within the company ; corresponding to asset_name in Asset Location table, trailer_number in Inspections table
Agency	Categorical	The name of tolls agency
Entry_Plaza	Text	Plaza name with identifier where the vehicle entered
Entry_Date	Time	The date and time the vehicle entered the plaza
Exit_Plaza	Text	Plaza name with identifier where the vehicle exited

Exit_Date	Time	The date and time the vehicle exited the plaza
Class	Categorical	
Miles	N/A	
Toll_Charge	Numeric	The amount of tolls
Dispute_Status	N/A	
Dispute_Reason	N/A	
Account	Categorical	

METHODOLOGY [15 points]

Describe your experimental design.

Predictive modeling-type project

- If you are doing predictive modeling, what sort of cross-validation are you doing and why?
- How did you partition data into training and test sets (e.g. 70/30?)? Why?
- What are you statistical and/or business performance measures? You might decide there are more than one to consider, such as AUC, Accuracy, Sensitivity, RMSE, MAPE, etc. Why are the ones you are using the most appropriate ones for your problem?
- Provide a diagram created in PowerPoint or Visio that shows the steps you took (e.g., queried data from DB, created new features, pre-processed data (how?), partitioned data, built model, evaluated models, etc.). This should make it crystal clear to the reader what your entire workflow does and help you explain to others in a PowerPoint or poster presentation later-on. Detail here is a good thing. Below is an example. Try to avoid using many or bright colors. Keep it professional and scientific.

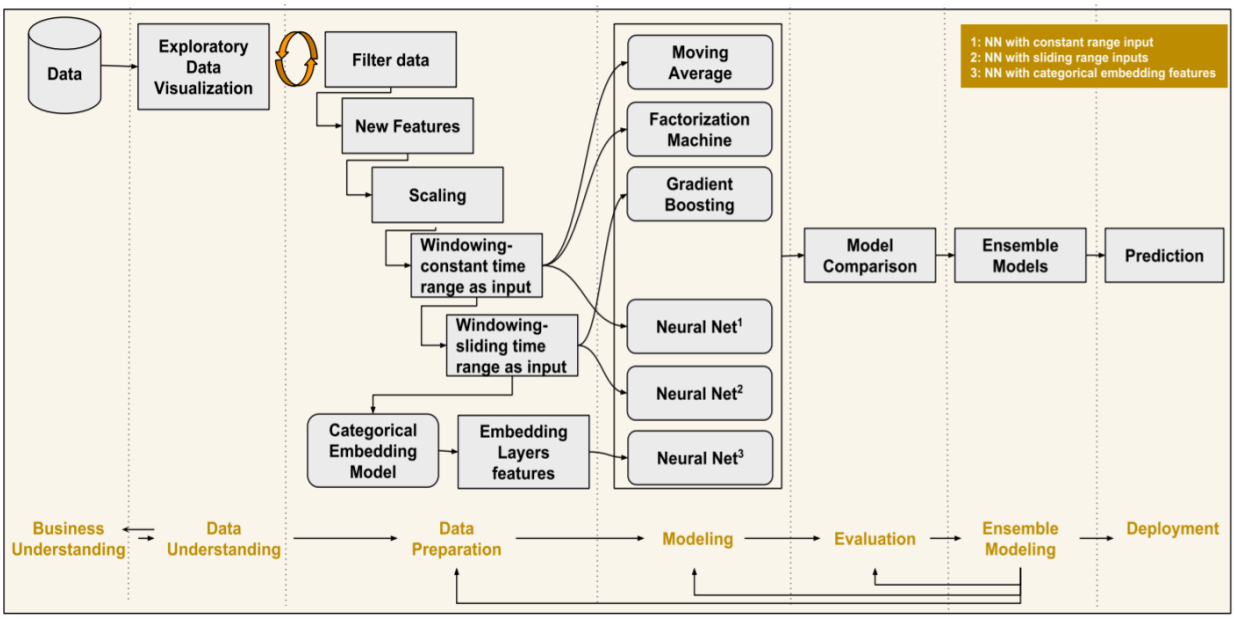


Figure 1: Analytics Workflow

Optimization modeling-type project

- If you are doing optimization, or there is a math model component to your project, diagram how the parameters (could be demand forecasts) are input as parameters to your model. Sometimes a process flow diagram is useful here to show the order of a process, when decisions could be made, etc.

MODEL(s) [10 points]

Describe/explain the models you used in your project.

Examples

- Neural networks are (in 4 to 5 sentences) … Pros/cons... Tuning parameters …
- Classification trees are (in 4 to 5 sentences) … Pros/cons…Tuning parameters …
- You will do this for every method you tried in your study.

Why did you choose to use these? Are there good reasons why they are relevant to your problem? Maybe you are using them because other studies have tried them as you indicated in your literature review.

Most likely you used various methodologies, or each model had various tuning parameters. Formulate each model using Overleaf or Word's Equation Editor. Be concise here. No need to write an entire page about a commonly known method.

If you are proposing a new or modified methodology describe how it is similar to other methods and what makes it different. For example, you learned how you can tweak OLS and make it a Lasso model. Maybe you tried to create/tweak a common algorithm to see if it improves performance. Use Overleaf or Equation Editor to put detail the method rigorously`.

If you are doing an optimization-type project, define your parameters, decision variables, objective function, and constraints.

RESULTS [15 points]

Summarize the results so that every model can be compared on the performance measures you found were important. I suggest to have both training and test statistics so you can make an argument if any model overfit the data or not. Here are a couple examples…

Table1. Average MAE (across same 160 time-series)

Model	Croston	Aggragate (NN)	GBM	NN	QRF	Meta (RF-QRF)
Train	33.51	30.45	26.31	28.11	29.03	28.88
Test	31.85	30.91	31.93	31.62	30.66	30.68

Classification Algorithm	SMOTE Ratio	Class Weight	Class	Precision	Recall	F1
Random Forest Classifier	None	None	Cancelled	0.96	0.58	0.72
			Closed	0.80	0.96	0.87
			Declined	0.58	0.20	0.30
Random Forest Classifier	Auto	None	Cancelled	0.86	0.58	0.69
			Closed	0.81	0.89	0.85
			Declined	0.44	0.29	0.35
Random Forest Classifier	Minority	None	Cancelled	0.87	0.58	0.70
			Closed	0.79	0.96	0.87
			Declined	0.54	0.17	0.26
Random Forest Classifier	Custom	None	Cancelled	0.91	0.58	0.71
			Closed	0.80	0.91	0.85
			Declined	0.46	0.27	0.34
Random Forest Classifier	None	Balanced	Cancelled	0.88	0.58	0.70
			Closed	0.80	0.91	0.85
			Declined	0.44	0.25	0.32
Random Forest Classifier	None	Custom	Cancelled	0.89	0.58	0.70
			Closed	0.80	0.91	0.85
			Declined	0.45	0.25	0.32
Random Forest Classifier	Auto	Balanced	Cancelled	0.84	0.58	0.69
			Closed	0.81	0.88	0.84
			Declined	0.42	0.31	0.35
Random Forest Classifier	Auto	Custom	Cancelled	0.82	0.58	0.68
			Closed	0.81	0.87	0.84
			Declined	0.41	0.30	0.35
Gradient Boosting Classifier	None	None	Cancelled	0.96	0.58	0.72
			Closed	0.84	0.95	0.89
			Declined	0.69	0.42	0.52
Gradient Boosting Classifier	Auto	None	Cancelled	0.94	0.55	0.69
			Closed	0.83	0.94	0.88
			Declined	0.63	0.37	0.47
Gradient Boosting Classifier	Custom	None	Cancelled	0.96	0.56	0.71
			Closed	0.84	0.94	0.89
			Declined	0.66	0.40	0.50

Table 7.2: Precision and Recall scores of the models

Use as many tables or plots of your results you feel are needed to describe your results to your audience, and identify anything interesting to the reader. A couple examples below…

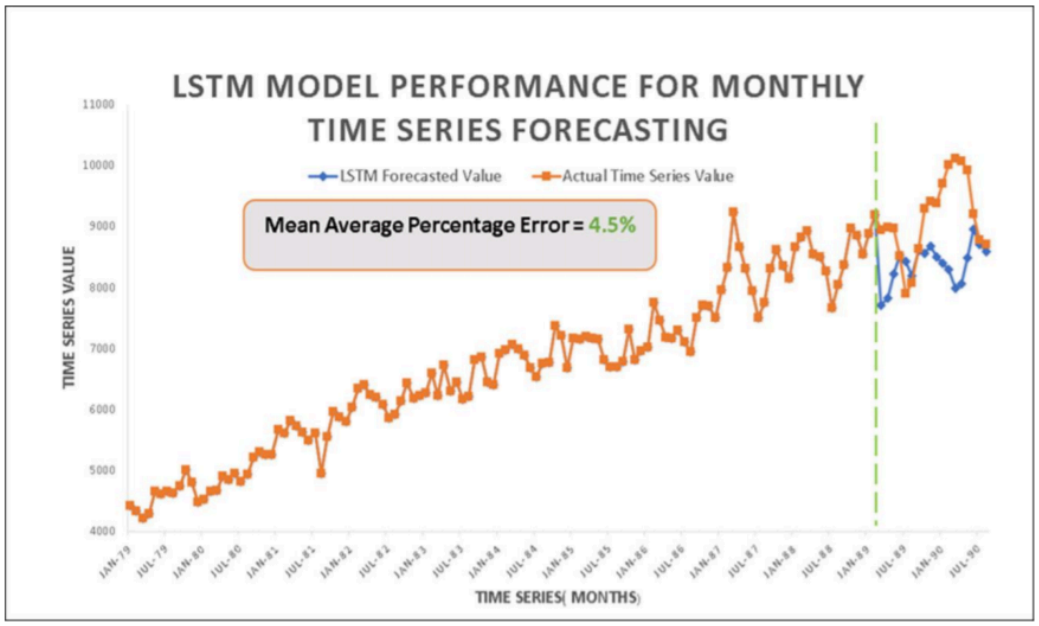


Figure 8: Plot of monthly aggregated time series and forecast

Classification Technique	SMOTE	Class Weight	Cost Saving Per Project
Random Forest	None	None	$30.99
Random Forest	Auto	None	$32.22
Random Forest	Minority	None	$29.53
Random Forest	Custom	None	$31.97
Random Forest	None	Balanced	$32.39
Random Forest	None	Custom	$32.22
Random Forest	Auto	Balanced	$32.61
Random Forest	Auto	Custom	$32.60
Gradient Boosting	None	None	$35.44
Gradient Boosting	Auto	None	$32.44
Gradient Boosting	Custom	None	$34.09

Table 7.3: Comparison of models based on cost savings per project

Discuss which model performed the best or should be used for decision-support and why.

Discuss any knowledge learned here that might be important to a decision-maker reading this section. Would a business want to use this best model to support their business? What can they expect from it?

CONCLUSIONS [2.5 points]

Restate the business problem and its importance.

Provide answers to your research questions.

Without restating everything you found in the results, discuss how your solution could be beneficial to the business.

What potentially strong assumptions did you make or limitations exist in your study that you believe are important for others regarding your solution?

What do you think requires more investigation to help support this problem? Usually after you answer your research questions, you likely came up with some questions to investigate that might make your models/solution better. The work is never done 😊.

REFERENCES

I encourage you to use EndNote or other tool when you cite your references to keep them organized and easily updatable. You can get a free account here (http://guides.lib.purdue.edu/EndNoteBasic). Make sure you cite all papers, codes, etc. appropriately.

APPENDIX

You can put supporting things here. Sometimes people will put their data dictionary table here and refer to it in the Data section.

INTRODUCTION

~~The logistics and transportation industry is undergoing a significant transformation, driven by advancements in technology and data analytics. One critical aspect of this transformation is the need for greater cost transparency and efficiency in fleet management. The Trailer as a Service (TaaS) program, which aims to provide flexible and cost-effective trailer solutions for businesses, faces significant challenges due to its reliance on manual processes to track and assign costs such as tolls, damages, and maintenance. These inefficiencies often lead to inaccurate billing, poor cost monitoring, and operational oversights, resulting in financial disputes and reduced transparency. The growing need to address these challenges highlights the importance of research into more efficient methods of cost tracking and operational management within the fleet management and logistics industry.~~

~~The significance of this issue is underscored by recent developments in the field. According to a report by Gartner (2023), businesses that adopt data-driven automation in fleet management can achieve substantial cost savings and operational improvements. In particular, automation enables companies to reduce human error, streamline billing processes, and enhance predictive capabilities in maintenance and asset management. Forbes (2022) further emphasizes that digital solutions in logistics are revolutionizing the way companies track and manage assets, ensuring greater accountability and minimizing unexpected operational costs. Additionally, an article in the Wall Street Journal (2023) highlights that organizations in logistics that integrate predictive analytics into their fleet management systems are better positioned to foresee potential issues and reduce operational disruptions, ultimately leading to a more efficient and cost-effective business model.~~

~~Given these trends, it is clear that the reliance on manual and outdated systems in programs like TaaS is no longer sustainable. To remain competitive and ensure long-term profitability, businesses must adopt innovative solutions that integrate automation, predictive analytics, and digital billing. These technologies can play a pivotal role in mitigating inefficiencies, ensuring~~

~~accurate cost tracking, and improving financial transparency. This study will focus on identifying and testing technological solutions that can optimize cost tracking processes for the TaaS program. The goal is to explore how automated tracking systems, predictive maintenance tools, and digital billing platforms can address the key challenges of inaccurate billing, maintenance oversight, and untracked operational costs.~~

~~The research question driving this study is: *How can innovative technological solutions, such as automated tracking systems and predictive analytics, improve the accuracy and transparency of cost tracking within the TaaS program?* This question is important because accurate cost tracking and financial transparency are essential for maintaining trust between service providers and their clients. Additionally, improving operational efficiency in the tracking of damages, tolls, and maintenance can provide a competitive edge in a rapidly evolving industry.~~

~~This research will employ an analytical approach that combines the evaluation of existing literature on predictive analytics, automated systems, and fleet management with an examination of practical case studies from industry leaders. By reviewing current practices, identifying gaps in existing systems, and testing potential solutions, this paper aims to propose actionable recommendations for enhancing cost transparency and operational efficiency within the TaaS program.~~

~~This study aims to provide valuable insights into the role of automation and predictive analytics in improving cost transparency and operational efficiency within the TaaS program, which could ultimately contribute to better financial accountability and greater success for businesses in the logistics industry.~~

This section should be at least 600 words.

Motivate why what you are you doing is important. Describe the business problem (without using the client's name if collaborating with an industry partner). Use at least 3 references in the media (e.g., Gartner, WSJ, and Forbes are good examples) that have been published within the last two years to motivate the problem and need for more research.

State your research question(s) clearly and explain why we should care about answering them.

In the process you should describe what you are you trying to do to support the business problem? In other words, describe big picture how your analytical approach will address the business problem needs.

Provide a brief preview of your argument and conclusions, then provide a roadmap through the paper like so in the in the last paragraph in this section.

> Example: "The remainder of this paper is organized as follows: A review on the literature on various criteria and methods used for supplier selection is presented in the next section. In Section 3 the proposed methodology is presented, and the criteria formulation is discussed. In Section 4 various models are formulated and tested. Section 5 outlines the performance of our models. Section 6 concludes the paper with a discussion of the implications of this study, future research directions, and concluding remarks."

(Version of introduction :

The transportation and logistics industry has experienced a major change because of modern technology alongside operational efficiency requirements. Trailers-as-a-Service (TaaS) model represents a new emerging trend in this sector which enables users to access trailers through a service-based approach instead of traditional ownership or leasing options. This cutting-edge service model creates new problems for device toll recognition and equipment condition monitoring functions.

The Gartner report indicates supply chain organizations will increase their investments in applications which enable artificial intelligence and advanced analytics by 2025 to reach more than 50% coverage. The logistics industry is now increasingly focused on making decisions based on data as research shows this emerging practice. Using current technological developments for toll attribution and damage tracking problems in TaaS operations enhances operational efficiency alongside customer satisfaction rates.

The Wall Street Journal reveals that trucking companies in America need to minimize their spending and enhance operational visibility according to new reports. Cost allocation and damage identification management in trailer operations requires absolute accuracy because it determines business profitability along with customer trust maintenance. Even though TaaS provides adaptability it raises difficulties for cost monitoring and which party bears what expenses.

Forbes predicts that Internet of Things-connected devices combined with advanced analytics technologies in transportation will generate $1.2 trillion economic benefits through 2026 according to published reports3. The calculated forecast highlights how technological solutions hold potential in resolving multiple industry problems which affect TaaS providers.

Our research targets two essential problems because of these circumstances.

1. What approaches from predictive modeling along with data analytics systems should be adopted to enhance both accuracy and efficiency of toll attribution operations within TaaS systems?

2.Establish methods that enhance trailer damage monitoring together with attribution processes despite existing inspection irregularities and trailer exchange frequency.

The identified questions require immediate solutions because they serve several vital purposes. Tolls should be accurately attributed to bill customers fairly while building trust between TaaS providers and their customers. An efficient system for tracking and attributing damage leads to decreased conflicts and it simultaneously enhances trailer maintenance operations and lifecycle management performance. Services that solve these problems will gain a strong competitive position in the modern logistics sector which is currently undergoing rapid change. The analysis will build predictive tools to automate toll expense allocation and damage inspection functions that support operational requirements of the business. Our research will review different techniques to pair toll payments with proper customers despite restricted data availability through the combination of license plate recordings timestamp records and GPS data. An algorithm needs to be developed to process large datasets effectively throughout making correct predictions with limited available information. Damage tracking and attribution will be handled through a multi-dimensional analytical model which consolidates data from inspections reports

together with telematics information together with historical accident records. Our research investigates the application of computer vision technology for pre-trip and post-trip inspection photo comparison automation which helps detect possible manual inspection misses of vehicle damages.

The following paper structure includes: Section 2 contains a literature review about transportation industry predictive models alongside damage detection techniques and toll attribution methods. Our proposed methodology structure comprises data preprocessing alongside feature engineering and model selection which appears in Section 3. Our experimental results with model descriptions appear in Section 4 following our method implementation description. The evaluation of our models' performance and practical applications and possible enhancement opportunities takes place in Section 5. Section 6 delivers the paper's summary including our study findings along with their effects on TaaS business and prospective research paths. The research we conduct addresses crucial operational challenges of the TaaS model with the goal to boost digital transformation within the transportation and logistics industry through efficient and transparent operations.)

Literature Review

~~(Predictive analytics is increasingly being used in the logistics industry to address challenges related to cost management, damage tracking, and customer satisfaction. By integrating advanced data models and leveraging telematics, predictive systems enable more efficient operations and decision-making. For a "Trailer as a Service" (TaaS) program, these technologies offer specific opportunities to enhance cost attribution, reduce inefficiencies, and improve billing transparency.~~

~~**1. Predictive Analytics for Cost Attribution** Cost attribution in logistics is a critical challenge, especially when multiple stakeholders—such as shippers, brokers, and carriers—are involved, as in a TaaS model. Predictive models can assign costs like tolls, damages, and maintenance to specific drivers, lanes, or trailers based on historical data patterns and telematics inputs. Studies ([TBU]) demonstrate that integrating predictive analytics with operational data can reduce disputes and improve cost transparency, enabling better client relationships.~~

~~This project aims to address the complexity introduced by brokers withholding data, making predictive models essential to bridge these information gaps. By triangulating available data from toll records, pre- and post-inspection reports, and telematics, predictive analytics systems can create actionable insights for cost allocation.~~

~~**2. AI for Damage Assessment** The use of AI-powered image recognition for trailer damage assessment is another well-documented application. The plan to automate damage evaluation through AI aligns with industry best practices, where image-based systems reduce manual workloads and improve assessment accuracy. Research by Park et al. (2021) highlights that AI can cut inspection times by half while maintaining high accuracy, thereby streamlining processes and reducing operational delays.~~

~~**3. Real-Time Analytics and Operational Automation** Real-time data processing is critical for predictive analytics to provide actionable insights promptly. In this context, this would involve real-time cost and damage attribution to prevent operational bottlenecks. Real-time toll analysis~~

and damage tracking can enable managers to address issues before they escalate. However, implementing such systems at scale remains a challenge due to data latency, integration issues, and computational demands.

4. Customer Experience Enhancement A key benefit of predictive systems is improving customer experience through accurate billing and transparent operations. For this project, clear cost attribution and reduced billing disputes can build trust and enhance customer satisfaction. Smith et al. (2019) emphasize that customers value transparent processes, which ultimately lead to stronger business relationships and loyalty.

Gaps in Knowledge and Areas for Further Research

1. Data Integration and Accessibility One of the critical challenges in the TaaS model is integrating diverse data sources such as telematics, toll records, and inspection reports. Questions remain about how to unify these datasets into a cohesive platform. For instance:

- What frameworks can enable real-time data integration?
- How can incomplete or missing data (e.g., driver details) be accounted for in predictive models?

2. Real-Time Cost Attribution The project underscores the importance of real-time cost attribution to specific drivers or trailers. However, implementing such systems requires addressing several unknowns:

- How should cost attribution criteria be defined, particularly for damages that cannot be directly associated with a specific driver or trip?
- How can predictive systems account for "edge cases" where data is unavailable or ambiguous?

3. Balancing Automation with Prediction While automation can improve operational efficiency (e.g., AI-based damage detection), predictive systems aim to forecast and prevent issues. Determining the right balance between these two approaches is an open area of exploration. For example:

- Should predictive models prioritize damage prevention over detection?
- How can automation tools be integrated into a predictive framework to maximize their effectiveness?

4. Cost Absorption vs. Client Billing A critical business question is whether to absorb certain costs or charge them back to clients. This decision requires clear, data-driven criteria, which are currently undefined. Research could explore:

- What metrics should determine cost absorption versus chargebacks?
- How can predictive models help inform these decisions in real time?

5. Ethical Considerations in Predictive Analytics Cost attribution systems must be transparent and fair to maintain customer trust. For this project, ensuring that predictive models are interpretable and explainable is essential. Future research should investigate:

- How to design predictive models that align with ethical considerations.
- How to communicate predictive outcomes to clients in a way that builds trust.

This predictive analytics initiative for its TaaS program represents a significant opportunity to modernize its operations, reduce inefficiencies, and enhance customer satisfaction. While there is substantial knowledge about the benefits of predictive analytics and AI in logistics, challenges such as data integration, real-time processing, and cost attribution remain. Addressing these gaps will require innovative solutions and further research, ensuring that the systems are not only efficient but also scalable and customer-focused.)

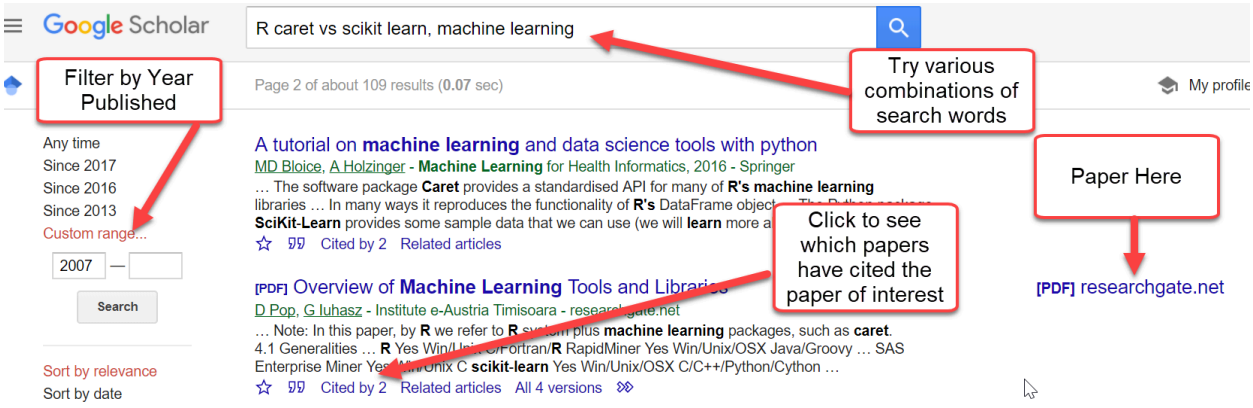
This section should be at least 700 words.

Here you are trying to summarize

1) what is known about the focus area you are working on, and more importantly
2) what is still unknown and thus necessitates further investigation.

To accomplish this use 5-10 academic journal references. I recommend you do the following: As a team, use Google Scholar and/or library databases to search for key words on what your study is about. Search articles not older than ten years old. Just using Google Scholar you will likely get several articles returned after filtering by year. Try to find one or two that are clearly what you are focused on. Ignore the ones that are "*sort of related*." Once you find those two highly relevant articles, you can click and see what other articles have cited them to see if you find even newer articles that talk about what you are doing. Download those handful of articles, then go directly to their references (listing within the paper at the end) to see if you see any other articles they have cited that are highly related to your study. You can search for them in Google Scholar to get them. At this point you will likely have more than 10 articles.

You goal is to find those one or two highly relevant and recent articles that give you most of the info you need (e.g. previously writing important articles to add to YOUR lit review). Once you find the golden nugget article, you are in great shape.



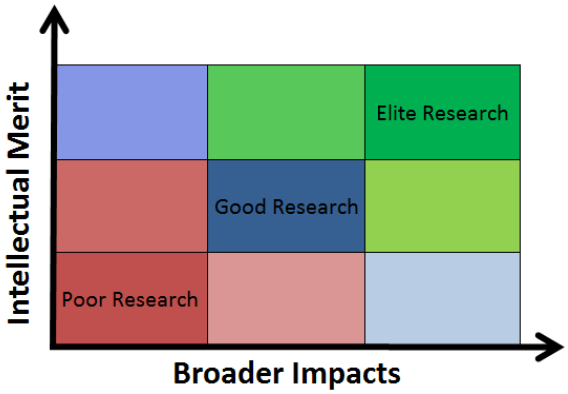
Now, based on the article titles, rank them based on the ones you think are most related to your work. Then one by one just read the paper **abstracts** to see if you think they are truly relevant to your study. If not, go to the next one.

Once you have filtered down the list, you are NOT going to read all these articles. That's right, you are NOT going to read these end-to-end. You will not have the time. What you are going to do is split these papers up among your teammates and each skim through the paper usually reading the introduction section to get some motivation for your introduction section, but most importantly identify what contributions their paper provides (**What** is novel about what they did?, **How** did they do it?). This could

be an approach, a model, a new method, new theory, etc. You should have a general idea based on what they said in the abstract. For each paper, you just need to write a few sentences that describe what their contribution was in the literature review section.

Next, as a team you are going to summarize what all these papers provided. Usually you can organize these contributions in the form of a table to make it easier to see for the reader.
Look here for many examples from previous MWDSI conferences: http://matthewalanham.com/

Lastly, usually at the end in the conclusions or discussion section of journal articles, the authors will mention items that are still not fully understood that still require more research/investigation. They might provide suggestions for future research. You will want to summarize this in two to three sentences. Hopefully, one of the papers that your teammates review identifies what you are doing as something that is important necessitating future research. This provides support for the value of what you are doing – it is support that what you are doing is NOVEL. If you cannot find such a case from those studies, you need to make a claim of why what you are doing is indeed important, novel, useful to a company, has implications to the field, could be extended to other areas, etc. Researchers will make sure their paper has both (1) technical merit and (2) broader impacts. For your project paper, I only expect broader impacts, meaning your work is actually useful at addressing some problem. Employers love to see professional publications from conference proceedings or journals on your resume. Journal articles are harder to achieve because you usually need to have significant mathematical rigor with proofs. Your potential boss/employer will not care about that, rather they will enjoy seeing how you framed a problem and logically developed and evaluated your solution. Thus, your project papers are perfect conference proceedings papers. However, some of you will be close to journal quality, and I will try to help you get that submitted to the appropriate journal if I feel we really have identified something more than broader impact that has intellectual merit.



Modeling-type papers on common public datasets (e.g., UCI Machine Learning Repo)

- If you are doing a project where your goal is to develop a better model like Author/Paper #1 used linear regression and obtained an R-squared of 0.78, Author/Paper #2 used decision trees and obtained an RMSE of 112.3, etc. Then you will need to record those numbers, which are usually found in the modeling or results sections of those papers. When you build similar models, you will want to have some statistical comparison of your models to others for comparison. This is a must if you are using the same dataset as others have used.

Modeling-type papers on non-common public datasets (e.g., Kaggle.com)

- If you are using data on a very recent Kaggle competition, there will likely not be any published papers using that dataset because it is too new. If the Kaggle competition is a forecasting problem, you just need some examples of expected statistical performance and ensure you are calculating those measures for comparison. Some might use stats like MAPE, or in classification F1 score, etc.

Modeling-type papers using proprietary datasets

- If you are using client's proprietary data, you will still want to have to have some stats from other studies to give an idea of what performance could be expected modeling in this problem area (e.g., supply chain, sparse demand, grocery forecasting, etc.). Also, if other authors have used a particular approach (e.g. logistic regression), you should also build such a model as a baseline for comparison. Maybe another paper used CART, then you should use CART and capture those stats. Hopefully you can make a case for using a different method to support the problem and compare those stats to the methods that have been tried in the past.

Non-business problem focused papers

- In the above examples, the focus is to create models that would yield better decision support than previously published models or approaches. However, you might be doing a project that focuses on creating a new algorithm or method, comparing some software packages, proposing a new work flow design for certain data science scenarios, or investigating some aspect of data mining. In these more classical research cases, we will work together to ensure all items required are covered. In other words, discuss with me, because every case is different.

At the end of the day, you are using these articles to provide you some guidance and ideas on how you can approach your problem. You can use these approaches as baseline models/solutions, which might work great for the industry partners problem (or class project problem). Most likely you will have some other ideas of your own, or there are other things that need to be considered based on partner feedback, etc. and this is where you can test or compare your model/solution to what others have tried.

A good literature review will convince the reader that your project fits into an established body of work and addresses a question of concern to the scholarly (or business) community, also that your project adds to our understanding of the topic by offering novel insights.