SMU Data Science Review

Volume 6
Number 2 *Summer 2022*

Article 19

2022

Analysis of First-Time Completion in the Field Service Environment

Gavin Rick
Southern Methodist University, grick@smu.edu

Scott Englerth
Southern Methodist University, senglerth@smu.edu

Marc Carter
Southern Methodist University, marcc715@gmail.com

Hayley Horn
Southern Methodist University, hayleyhorn@gmail.com

Analysis of First-Time Completion in the Field Service Environment

Marc Carter[1], Scott Englerth[1], Gavin Rick[1], Hayley Horn[2]
Master of Science in Data Science,
Southern Methodist University, Dallas, TX 75275 USA
{marcc, senglerth, grick}@smu.edu
2 Sr. Solutions Architect, Databricks,
160 Spear St 15th floor, San Francisco, CA 94105 USA
hhorn@smu.edu

Abstract. First-time completion is an important measure of service quality and efficiency in the field service industry. Customers call upon field service providers to repair their equipment in a timely manner so it can be put back into service for their business demands. Responsiveness can be measured through first-time completion and is defined as completing the repair on the first visit of a service call. This research is exploring the first-time completion in the forklift service industry. This research found the primary factors that impact first-time completion percentage in this industry include parts on hand, parts backorder process, technician experience, and anticipating service demands.

1 Introduction

The field service industry, and specifically the forklift service industry, strives to achieve 100% first-time completion which means increased efficiency, revenue, and customer satisfaction for the servicing dealership. While the goal is to resolve all service calls on the first visit most dealerships struggle with first-time completion. They also do not fully understand the factors that affect their ability to complete a job the first time. The objective is to improve service efficiency through first-time completion utilizing Machine Learning (ML) optimization models as well as processing previously unused technician write-ups into useful data through Natural Language Processing (NLP). In a field service environment, first-time completion is an important measure of efficiency for a dealership. First-time completion commences when a customer submits a service request, a technician is dispatched to the customer's site and completes the repair on the equipment the day they arrive. They will ensure the forklift is operating safely and correctly without having to leave to get parts, tools, or another technician. First-time completion is essential to the service dealer as higher first-time completion allows for more service requests completed per tech per day. It is equally beneficial to the customer who values a quick and efficient repair so their equipment can be returned to service as fast as possible. Typically, the customer's expectations are met if the servicing dealer can achieve first-time completion on the service request. If a customer's equipment is down too long, the technician does not arrive on time, or the repairs cannot be completed quickly and efficiently, then the customer might call upon a competitor to service the equipment. Perceived fairness, empathy, responsiveness, reliability, and convenience all affect the customer's service quality satisfaction (Andaleeb, S. S., & Basu, 1994). Unsatisfied customers that experience excessive downtime could result in a loss of trust and revenue for the dealership. Keeping the customers' equipment in service will increase customer satisfaction, gain confidence, and cultivate the relationship. Even a slight increase in client satisfaction of 5% can lead to a long-term profit increase of 25-85% in profits (Reichheld F. F., Sasser W. E. 1990).

First-time completion can be calculated in a multitude of ways, spanning from quite simple to extremely complex. The easiest way to capture first-time completion is to allow the technician to designate the repair as first-time completion. However, this could allow for incorrect designations and should be used with care. The next method is more automated and inspects the labor postings of the technician to ensure only one labor posting has been applied to the repair order. This method is a more reliable option and will be used for the purposes of this paper. There are many other ways that service companies might measure first-time completion; however, they all condense down to answering the question, "Did the technician complete the repair safely and correctly with only one visit to the customer?"

The success of field service first-time completion is impacted by numerous factors (Dutta 2013). Parts availability, technician experience, and additional unexpected repairs are just a few factors. As with most industries, the pandemic in 2020 and 2021 impacted the supply chain for parts and technician talent needed to make the repairs. In the service industry, parts availability is a multi-faceted problem. Ideally, the servicing technician would know the service required beforehand and arrive at the customer's location with the needed replacement parts. However, this is rarely the case as the technician typically only knows the customer has a service request for one of their forklifts and little else. Usually, the repair needs are unknown or unclear to the technician until he arrives at the customer's location and assesses the equipment. Even if the technician comes prepared for a detailed repair request, oftentimes there are additional unexpected repair needs he discovers upon inspection. As the technician discovers additional needed repairs the steps to complete the service are as follows:

> 1) Ideally the technician carries the parts needed for the unexpected repair on his van. The parts he carries on his van are limited by van capacity and rely on each technician's anticipation of repair needs in their territory as well as their own inventory management. If they do not have the part on their van, they proceed to the next option.
>
> 2) The technician makes a trip to the nearest shop, picks up the required parts, and returns to the customer site either on the same day or the next day, depending on timing. The success of this depends on how far the technician needs to travel to the shop and the parts availability at that location. If the part is not locally available, then the tech proceeds to the next option.
>
> 3) The technician calls the parts department to order the required parts and will return to the customer when the parts are available. Given post-pandemic supply chain issues this oftentimes severely delays the time between a

> customer's request for a repair and the repair being completed. All the while the customer's equipment is down, and they are not able to fully function at their warehouse further compounding supply chain issues in other industries.

At its core, this is an operations supply chain problem, with service van capacity and storage costs being the main constraints. Through machine learning, this study aims to better prepare technicians by anticipating repair needs and stocking their trucks accordingly based on the trucks in the technician's territory and expect repairs based on the characteristics of the customer's fleet of trucks.

2 Literature Review

2.1 Inventory Management

First-time completion is dependent on technicians having the right parts for the repair on their van or in a local parts consignment. The parts needed for the service job are often unknown to the technician before he arrives at the job. Managing their service van's capacity becomes a micro inventory management problem for each service technician. Research in this field has been extensive. Both academics and practitioners have investigated the components of inventory management strategies such as setting order policies, defining capacity constraints, and various methods of demand forecasting.

To best determine the service parts required to complete a job and whether a part is available one attempt was made to use machine learning to predict problematic parts that could be causing the service call. It was determined that problematic parts also impacted service quality and behaviors of the servicing technician. (Young-Hwan Choi, et al., 2022). Technicians realize the importance of knowing what part to have available at the time of the service call to increase customer satisfaction as well as reducing the number of visits to the customer location.

Some studies have addressed inventory management from a single item demands perspective (Sinha, 2014). However, the problem this study aims to address can be better defined as a multi-item problem as forklift parts are often ordered and then used for a repair in conjunction with one another. Industry educated parts purchasing policies for all stocked parts should be considered that focus on less inventory and a reduction in stocking the wrong or slow turning parts. When numerous items are ordered from the same vendor, certain fees that are charged for the order will depend on the number of ordered parts, weight, freight, and ordering frequency. Therefore, the order quantity and frequency of each item cannot be decided on demand (Nagasawa, et al., 2014). The service vans used to carry the forklift parts have a finite capacity and must be considered when stocking parts for repairs. Order and transportation costs change depending on the capacity of the van and the number of parts needed for repairs that are carried out. A dynamic order-up-to policy that provides full truckloads has been proposed by Kiesmüller (2009). Kiesmüller (2010) created approximation math models for calculating the factors of the replenishment policy. His models factor in the capacity restrictions of the total order amount size, so the target service levels can be met.

2.2 Technician Knowledge sharing to improve First-Time Completion

One issue that can occur during a repair is the technician does not have the skill set or knowledge to correctly analyze and diagnose the issues affecting the equipment. If the technician is unsure how to correct the problem this can have an impact on first-time completion percentage. One way to correct this is additional and higher quality training for the technicians. Proper training can be costly and more importantly time-consuming since they cannot bill their time spent training. On the other hand, insufficient training can lead to greater long-term costs due to long troubleshooting efforts or having to pull in more seasoned technicians off their jobs to assist. Sometimes on-the-job training is a fantastic way to learn but having two technicians for a single repair is not efficient. Technology can be used to fill the gaps. Using technology to connect the technician to an expert while in the field allows the technician to feel more competent, connected, better prepared, and attain a sense of achievement. (Kaasinen, et al., 2018**)** Being able to connect to an expert or team leader means that an inexperienced technician might be able to complete a repair in a single visit to the customer rather than having to send another more experienced technician to complete the repair a different day, which would fail the first-time completion metric. There are many technologies that could provide great assistance with training like cell phones to call another tech or tablets to look up manuals. However, a more robust technology that could help in this endeavor is augmented reality (AR). AR has successfully been used to have high potential in maintenance, training, repair, service, and inspection. While the potential for AR is high in these areas, the technology still needs to improve by engineering devices to be more portable, accurate, and stable as well as smaller, cheaper, and lighter. (Nee, A. Y. C., & Ong, S. K., 2013) A team leader or domain expert can provide knowledge through AR and develop newer technicians into fully productive and experienced technicians.

Technical Customer Services (TCS) plays a key role in the field service industry. Due to the large array of tasks, the ability for a technician to have immediate access to support is key to their success. Service technicians must deal with many tasks when onsite at the customer location. These include maintenance tasks such as safety inspections, customer support, and actual repair of the equipment. The technician must also master the operation of the equipment to ensure they know it is operating correctly. This fieldwork is dependent on current information about the standard repair work that is required for that equipment.

To assist with this wide range of tasks, the technician requires product knowledge, and sufficient training on his tools and equipment in his territory. To assist in his job and support him in his tasks a mobile software solution should be in place for his support. Due to the field service technician being mobile and needing to be at various places to service equipment a mobile device must be used. An efficient mobile solution should include enterprise data and how long a job should take to complete so he can compare his efficiency. Mobile solutions allow for flexibility and functionality for the technician to support the customer's needs and potentially increase productivity. (Matijacic, et al., 2013).

Utilizing technology such as a mobile solution can assist the technician with benefits such as service history on the equipment in the customer fleet of forklifts as well as the condition of the application of the forklift such as a freezer or other factors that could impact the performance of the equipment. The pyramid model was created by Parasuraman to show the impact of technology in the value of servicing customers. This model shows has technology can be used to improve overall customer experience. (Parasuraman, et al., 2000).

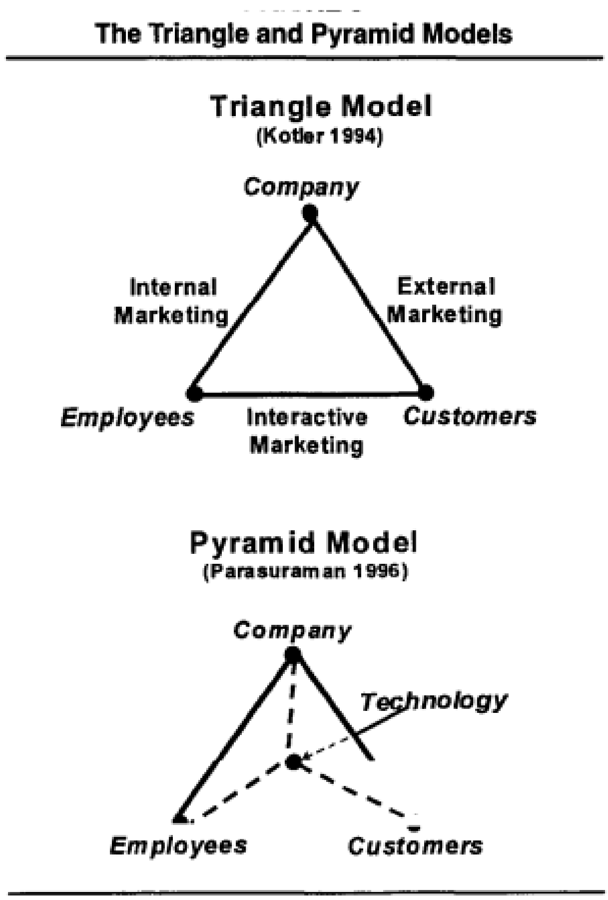

Figure 1. The Triangle and Pyramid Models

In another study linking technology and field service, it looked to manage problems such as first-time completion as the following "As business increasingly attempt to integrate all their function in a "supply-chain" framework, focus will shift from solving "independent" functional problems to designing and managing an integrated and interdependent enterprise." (Agnihothri, S, et al., 2002, p. 7).

Dispatch is also a key element in the field service industry and ensuring the technician is successful by using a few factors of assigning the right technician to the service call, that they are trained to handle, as well as meeting the customers' expectations of communication from the assigned technician. The dispatcher is also the first point of contact with the customer to help determine the purpose of the service call to better prepare the technician for the repair. The four key items a dispatcher must do while on the call with the customer are the following: (Blumberg, 1994).

1. Fully Identify the service call requirements to understand the service need
2. Evaluate the call to determine whether to dispatch a technician to their location.
3. Determine that based on previous calls and history of the customer, the customer clearly is explaining the service call issue to determine if a site visit is required.
4. Document the service requirements for the call and assign the appropriate technician with the skill level and parts available to complete the job.

While the dispatcher must balance all these needs as well as successfully plan the technicians' schedule it is another area of focus to help maintain first-time completion.

2.3 Technician Efficiency

Another impact on first-time completion is the experience of the technician and their ability to diagnose and effectively repair the equipment. To determine technician performance, previous service history needs to be analyzed and their effectiveness quantified. To evaluate technicians' impact, we review two different metrics including technician productivity and technician efficiency.

Technician productivity can be quantified by the following equation (Corrigan and Tsimpinos, 1994)

*technician productivity = labor hours sold ÷ operation hours * number of technicians*

Technician efficiency is measured by the following equation. Using this number to determine how efficient a technician is will show what technicians are more efficient than others and give a guideline to aim for in determining their effectiveness.

technician efficiency = labor hours sold ÷ hours worked (including over time)

Once the calculations have identified the areas that have the most impact on productivity, a performance audit can be applied as a tool to increase productivity and find areas of improvement. As a result, focus can be shifted to technical training to help bridge any knowledge gaps between technicians. Including auditing the following skills and areas for improvement. These include technical experience, mechanical experience, areas of current certification, and efficiency training attended by date to stay current. We also consider individual areas of expertise, cross training, tools owned and technical competence.

For the technician to have the highest grades the biggest factors that impact their performance are training and experience. Since experience comes with time and the number of service jobs, we can only focus on the training and education of the technician.

The easiest way to measure technical skills or a technician that needs additional training is to look at similar service work orders and the average resolution times per technician for each type of service. Technicians who take longer could use additional cross training on common service issues and their resolutions.

2.4 NLP to extract useful repair categories

Natural Language Processing (NLP) is a component of Artificial Intelligence (AI) in the field of linguistics that deals with the interpretation and manipulation of human speech or free text using software. NLP as a tool can be used to extract useful information from product reviews such as detecting sentiment or medical records for reviewing the treatments applied on the patient write-up and even automotive repair orders for the steps taken to resolve an issue. Unfortunately, existing commonly used data collections, such as the Penn TreeBank, are not effective at processing repair notes of technicians. Often the notes have specific verbiage associated with the service industry, are usually brief, and do not obey the standard spelling and grammar rules for English.

To extract value from the free-text technician, write-ups supervised, or unsupervised learning techniques can be utilized. A study by Khanbhai, Mustafa (Mustafa, 2020, p. 1) found "that comments extracted from social media were commonly analyzed using an unsupervised approach compared to free-text comments held within structured surveys were analyzed using a supervised approach" when comparing NLP methods applied to free-text data in the medical field. Furthermore, the study identified data quality, semantic and synthetic relationships in the data, and dataset size, as three of the key components that drive the decision making around which ML model to use.

A study by Solanga (Solanga, 2020, p. 41) on short comments in a Navy feedback survey used Frequency analysis, LDA (Linear Discriminant Analysis), Networks, an iterative procedure using CLS labels, as well as a web-based approach to create comment topic bins. They then used text processing techniques to apply labels to each comment and assign each labeled comment to a topic bin. Through this method of automation of binning topics, Solanga was able to also address the challenge of handling the leftover comment labels that are set into the other category. “By utilizing the next highest CLS labels and allowing the user to bin them according to a relevant topic, one can readily handle a significant amount of obscure comment labels that would have been previously left as "Other."”

Entity extraction on the technician notes attempts to process the technician's notes to extract searchable entities such as parts used, and labor completed. A research team used Hidden Markov Models (HMM) to process the data after parsing and extracting the noun phrases. (S. Bratus et al. 2011) The research for this project aims to categorize the repairs based on the extracted noun phrases and not necessarily develop a searchable taxonomy. Instead of using an HMM (Hidden Markov Models), this study will use a Random Forest model to develop the expected category of the repair. Once the category of the repair is identified it can then be utilized in other models to help determine and improve first-time completion.

To get useful information from frequent words and word combinations, proper word vectorization is of importance. A study by Mikolov et al. (2013) shows that a combination of simple vector addition and the representation of phrases with a single token can be an extremely efficient way to represent longer pieces of text.

Another obstacle for standard NLP pipelines may be the domain specific language used in the free-text responses written up by the technician. This is not a unique problem to this specific data. MaintNet is a communal open-source library of technical and industry specific natural language datasets, that provides logbook data from the airline, automotive, and facilities industries including the tools that aid in their (pre-)processing and clustering (Akhbardeh, et al., 2020).

3 Methods

3.1 Data

Data used in this study was supplied by a North American based forklift dealership, which provides sales, service, and parts to support its customers. This dealership not only maintains forklifts but also batteries, chargers, conveyors, doors, and other equipment. To determine factors that affect First-Time Completion (FTC) the dealership removed any repair work orders that did not pertain specifically to forklifts. As it is not expected that the dealership would stock the parts required for anything other than a forklift. Furthermore, PM (preventative maintenance) work orders have also been excluded from the datasets as it is expected that a PM work order would be completed on a single visit and artificially drive up the FTC metric the dealership tracks.

The dealership supplied a data spreadsheet in a CSV format, which contained data for each of the years 2019, 2020, and 2021. The data was summarized to the work order level and contained almost 315K rows and 26 features. A few notable features are the "top_notes", the technician's skill level on a specific piece of equipment, and the determination of first-time completion. The "top_notes" are the technician's write-up for the work order which could be a repair or a warranty work order and could provide great insight into why a repair was completed as FTC. Technician skill level is the second notable feature and could be established to be a dominant feature in determining FTC. This feature is specific to not only the technician but also the equipment attached to the work order. Another feature is the designation of FTC and is defined by this company as a repair that was able to be completed during a single visit to the customer.

3.2 EDA (Exploratory Data Analysis)

3.2.1 Missing Values

Some features contained null values, but we were able to fill these with sensible data based on subject matter expertise. Features such as: "action_code"," failure_code", and" equip_series" were all set to "NONE", if found to be null. A few exceptional cases did exist such as "years_exp" the decision was made to fill these blanks with the mean value from this feature. Another case is the feature "tech_skill_level" these features could have been set to "NONE"; however, it was felt that even if a technician has never seen a particular model of forklift before, they have seen similar forklifts and therefore have basic experience. With this logic, these blanks were set to "LOW" instead of "NONE".

First-Time Completion variable distribution is fundamental in determining the causes of missing this essential service metric. Seen in Figure 1 the distribution of this feature is skewed towards the positive value or FTC by more than double and could cause issues during modeling. Techniques such as oversampling and under sampling will be used to combat this issue of imbalance.

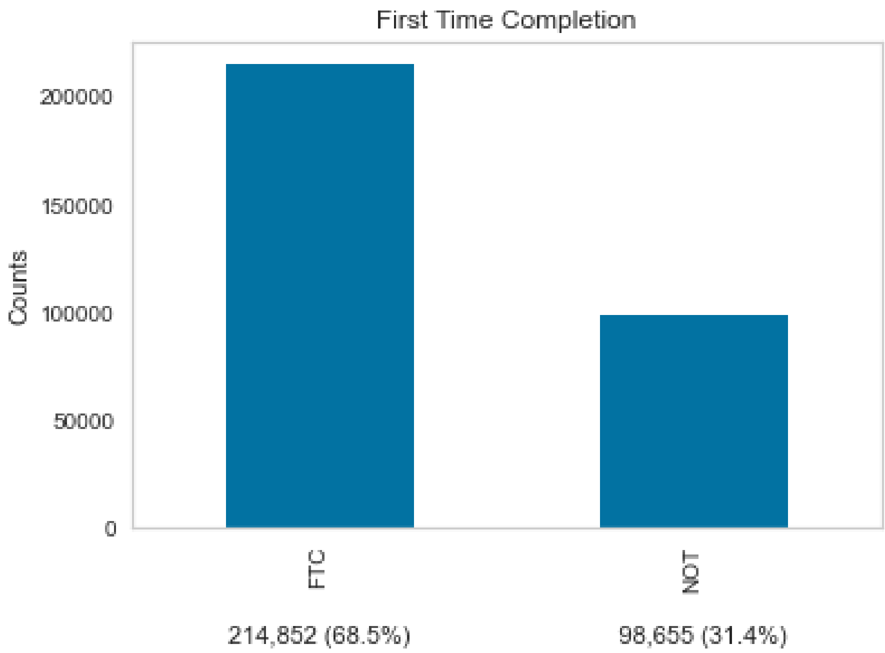

Figure 2. The distribution of First-Time Completion.

First-time completion by equipment group provides some interesting information, such as the "RCH" group has over double the successful first-time completions when compared to not achieving first-time completion. Other equipment groups also have high success rates, while some are almost equal like the "IC" group. A few equipment groups are upside-down in the number of successful first-time completions like "LIFT" and "SL", both groups need to be looked at closely to determine if a cause can be determined.

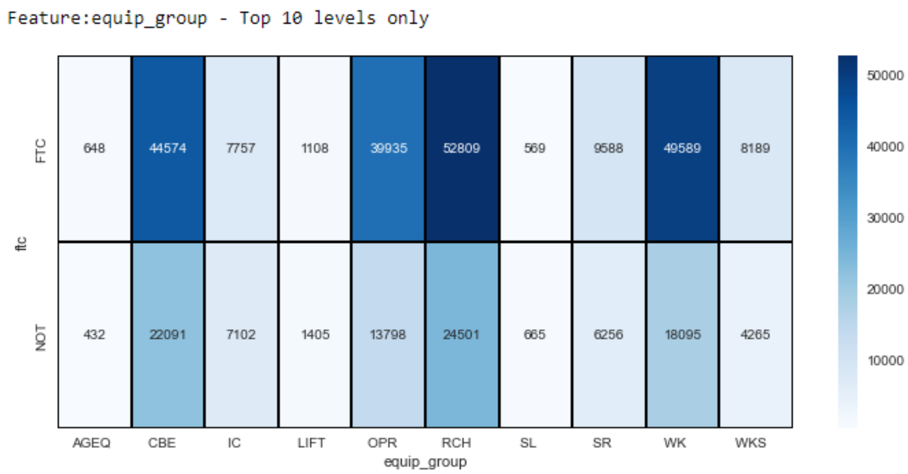

Figure 3. First-Time Completion by Equipment group.

3.2.2 Text Field Analysis

An integral part of this project is the analysis of the three free text response fields included in the data. The three fields available in the dataset were "Top Notes", "Bottom Notes" and "Instructions." Top Notes are the notes the technician makes while on the job about the repair, the “Bottom Notes” are the notes between the technician and dispatch, and the “Instructions” are notes between the technician and the warehouse. Through tokenization, word aggregation and visualization of the three columns this study was able to narrow down its focus on the “Top Notes” column as the main free-text column to use for analysis. The three main reasons for this choice were 1) this column had the most words when compared to the other two, 2) through tokenization and word cloud visualization this study discovered that the frequent words used in that column indicated its usefulness, and 3) the technician's comments while on the job describing the repair are likely to be the best source for data that will drive the analysis around first-time completion.

The number of words available to analyze is a principal factor in the usability of the free-text data. More meaningful words in the text allow for more prominent differentiation between data points and more defined text clusters. Analyzing the word count of the three data points did not only provide guidance as to the usability of each column for NLP, but also revealed a significant difference in text length between repair jobs that were first-time complete and those that were not.

```
Top Notes Word count: 44.18398117162505
First-time completetion Top Notes Word count: 33.957333159433006
Not First-time completetion Top Notes Word count: 57.40820582012204
Bottom Notes Word count: 8.4306675606942
Instructions Word count: 20.169298788780157
```

Figure 4. Mean word count of free-text response fields

The clean text, after the removal of stop words and punctuation, was visualized by a word cloud. The word cloud uncovered two things. First, the words in the word cloud are reflective of technician's problems and solutions on the job. This information should contain useful data points for first-time completion analysis. Furthermore, the theme of three topics within the “Top Notes” was revealed through the word cloud and the word count frequency analysis: Cause, concern, and correction. This finding was also re-affirmed through word vectorization and k-means clustering that revealed the three topics outlined above.

Figure 5. Top Notes Word count

3.3 Technician's Notes Clustering

During EDA three key words in the technician's top notes were discovered: "Concern," "Cause," and "Correction." The technician's top notes text field was then separated by the key words to create three new text columns. Each of the three columns was then vectorized and clustered through k-means. Clustering results were evaluated by word frequency in each of the clusters and domain knowledge. The emphasis was to create clusters that are granular enough to differentiate the biggest repair concerns, while not adding too much complexity by creating too many clusters. K=5 showed the best cluster results for each of the three key word columns. Cluster labeling was reliant on domain knowledge, but the focus of this exercise was to create useful input variables for further modeling.

The concern cluster was labeled as: “Leak,” “Wheels,” “Error Code,” “Battery,” and “Light.” The cause cluster was labeled as: “Wear&Tear,” “Damaged,” “Wheels,” “Electrical,” and “SM.” The correction cluster was labeled as: “Email,” “Wheels,” “Returns,” “Found,” “Removed.” Observations where Correction, Cause or Concern were missing were labeled as “None” and created a sixth cluster for each of the columns. This decision was made because these words missing from the technician’s write-up might be indicative of the repair character and not something that should be removed or imputed.

3.4 Data after pre-processing and EDA and one-hot encoding

After EDA, pre-processing, and text clustering the dataset contained a total of 24 features, 16 of which were categorical and 8 continuous. To prepare the categorical variables for modeling they were one-hot encoded, using the built-in “get_dummies” function in Python’s Pandas software library. One-hot encoding expanded the dataset from 24 features to 3317 features. The data was then split into training and test sets at an 80/20 split, using Sklearn’s “train_test_split” method with a random state of 55.

While some of the models inherently deal with class imbalance issues the lasso and logistic regression do not, therefore over-sampling was used to balance the training dataset. The technique used is called SMOTE (Synthetic Minority Oversampling Technique) and is implemented from “imblearn” and specifically the “over_sampling” method was used to balance the dataset. The decision to over-sample rather than under-sample was simple, due to the nature of the data and trying to determine the factors that are highly associated with FTC. In under-sampling the majority class has samples removed until it matches the size of the minority class, therefore highly valuable information from the majority class (in this case positive FTC class) would be lost.

3.5 Feature pre-selection through Lasso

L1 regularization (Lasso) in conjunction with linear or logistic regression is commonly applied to models to improve performance and prevent overfitting. The penalty term λ in L1 regularization is applied to the absolute value of the feature coefficients and can introduce sparsity. This means that as λ increases feature coefficients are reduced and can be reduced to zero. The goal of Lasso feature selection in this study was to reduce the dataset size for future modeling and improve the performance of H2O’s AutoML which will be discussed further below, while keeping the variables most important to first-time completion in the dataset. With that goal in mind, λ was not set to the value that produced the best predictive results, but to a value that was low enough to keep enough features in the dataset for efficient further modeling. At λ=0.001 the feature list of 3316 was reduced to 46 feature columns, using Sklearn's Lasso model.

In conjunction with models developed from H20 other models were developed and tested and for those models the best lambda (alpha in python) needed to be determined. Running a cross-validation model for Lasso regression and fitting the SMOTE dataset the best lambda was determined to be $9.4599e^{-5}$. This value of lambda selected 303 features for use in the other models and shrank 3014 features to zero. A new dataset containing only the selected features was developed and used for the logistic regression, random forest, and boosted tree model.

3.6 Logistic Regression

Logistic regression would be used to develop a baseline and was not expected to return remarkable results. The decision to use logistic regression as the baseline was due to the fact it is a good model for classification problems and is simplistic in its implementation. To determine the best parameters to use a cross-value grid search was used to look at combinations of the parameters. This produced a set of optimal parameters for the model as: (C=1, penalty=" l1", solver='liblinear', max_iter=1000). This model was able to achieve an accuracy score of 76.7% with a precision, recall, and AUC score of 77.9%, 36.7%, and 65.9% respectively. The output values can be seen below in figure 6.

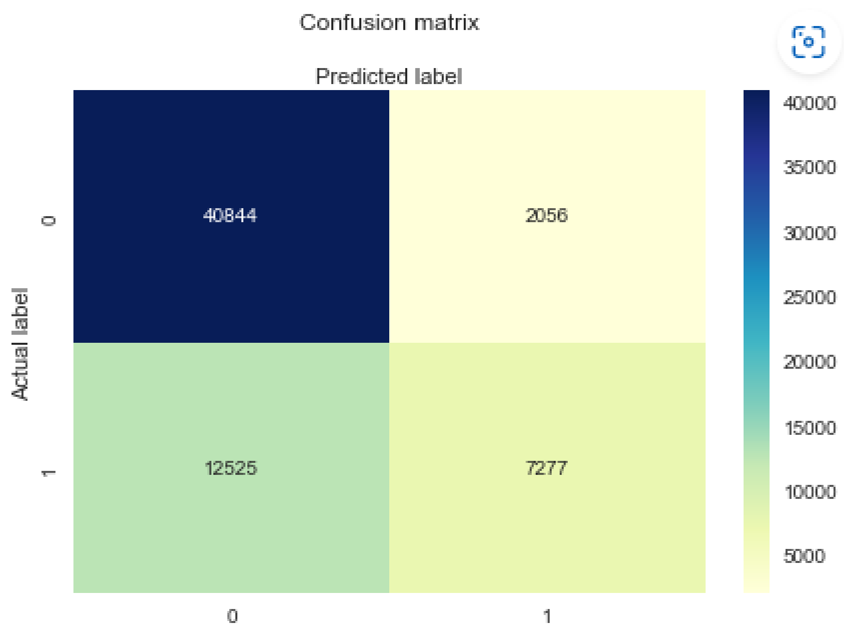

Figure 6. Logistic regression confusion matrix

3.7 Stacked Ensemble, XG Boost and GBM

In addition to the logistic regression baseline model, a few other models were developed such as a random forest and a boosted tree model but did only marginally better than the logistic regression model. Therefore, a new method for determining the best model had to be found. H20 was chosen because it is an open-source modeling software that compares many models at once.

The reduced dataset of 46 feature columns was ran through H20's AutoML API. Using H2O's AutoML machine learning workflows can be automated, including training and tuning of many models within a user-defined period. A key-feature of H2O's AutoML is that it provides a simpler wrapper function for many modeling related tasks that would normally require a lot of advanced code. The current version of AutoML trains and cross-validates through a variety of different models. For this study four gradient boosting machine models (GBM), a generalized linear model (GLM), three XG Boost GBMs, a default random forest model (DRF), and an extremely randomized random forest model (XRT) were applied. In addition, AutoML trains two Stacked Ensemble models. One is a "All Models Stacked Ensemble" which combines all the base models ran, and the "Best of Family Stacked Ensemble" which includes the best performing model in each family for each of the model types outlined above. In total this added up to 10 based models and 2 ensemble models being ran through H2O AutoML. In addition to defining the number of models ran this study also took advantage of balancing the dataset through H2O's pre-defined class balancing input to adjust for the class imbalance. As mentioned before, class balancing was done manually for the models that were not ran through AutoML.

4 Results

4.1 Measuring and comparing Model performance

To compare model performance across all models, this study used the AUC score to determine the best performing model. The AUC score reflects how well a model classifies each class in a binary classification model. Scores can range from 0 to 1. A score of 1 means that both classes are predicted perfectly, a score of 0.5 means that predictions are at random and as good as chance, and a score of 0 means that classes are predicted perfectly imperfect. In addition to the AUC score this study also used the common confusion matrix metrics such as Accuracy, Recall, Precision, and F-1 Score to compare the performance of the models. See Figure 7 below for a comparison of model performance metrics. For the GBM and XG Boost GBM models where multiple variations were ran, only the metrics of the best performing model are displayed.

Model Name	AUC	Accuracy	Pecision	Recall	F-1 Score
Logistic Regression	0.779	0.767	0.779	0.367	0.499
GBM	0.916727	0.8464	0.7984	0.6844	0.737
GLM	0.854767	0.7969	0.766	0.51	0.612
XG Boost GBM	0.913345	0.8448	0.791	0.6887	0.736
DRF	0.911812	0.843	0.777	0.7016	0.737
XRT	0.874033	0.799	0.759	0.5287	0.623
All Models Stacked Ensemble	0.917749	0.8485	0.7944	0.6969	0.742
Best of Family Stacked Ensemble	0.9176	0.8484	0.793	0.701	0.744

Figure 7. Model Performances

Model performance is important to evaluate which models perform well in classifying first-time completion. However, the goal of this study is not to predict first-time completion, but determine what factors drive first-time completion, meaning that feature importance is more impactful to the discussion.

4.2 Feature importance

The SHAP Summary plot below shows the contribution of the feature for each instance of data in the best performing tree-based model used in the All-Models Stack Ensemble model. Feature observations to the right show a strong predictive relationship between observation and first-time completion. Feature observations on the left show a strong predictive relation between the observation and non-first-time completion.

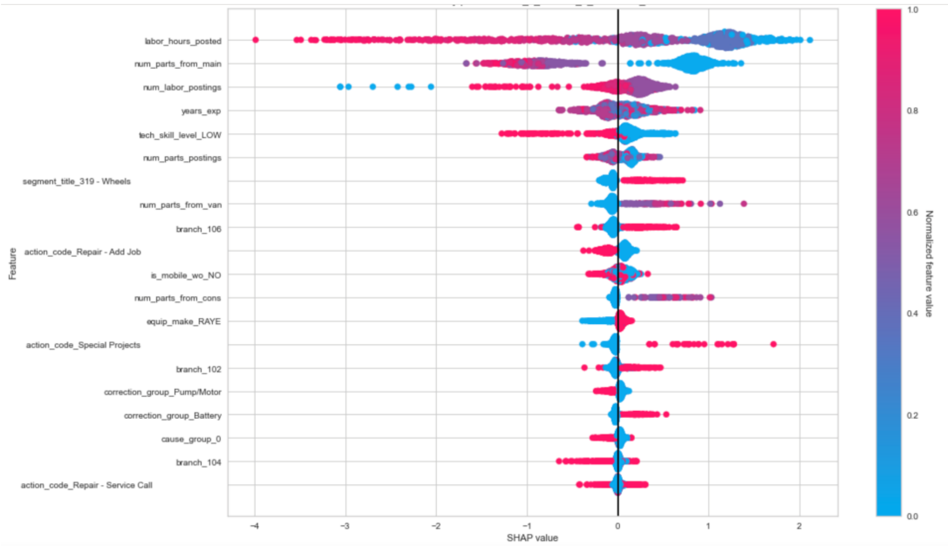

Figure 8. SHAP Summary

Some interesting insights emerge when looking at the SHAP summary chart. For instance, labor hours posted shows that the lower the number of hours posted to a work order the higher the percentage of FTC, while conversely the work orders with a higher number of hours usually means FTC was not met. But this feature needs to be looked at in conjunction with the number of labor postings feature. This feature shows that apart from a few outliers, which can be explained, the more labor postings a work order has the less likely it is to meet the FTC metric. Combined these features are suggesting that work orders with few postings and low hours are more likely to be considered FTC, which makes sense because the more hours needed to finish a job the more likely it is to need a second visit or parts not stocked on the van.

Another combination of features that provide great insights are the number of parts posted from main, van, consignment, and total parts posted. Inspecting each feature independently does provide some information but can be confusing. Take parts from consignment for instance, the chart would suggest that posting only a few parts from consignment would indicate a reduced percentage for FTC. This does not make sense, because the consignment holds universally used parts to the customer directly and are not posted from the consignment unless the consignment has the part in stock. But when grouping the parts features together and inspecting them as a group, it should become clear that if a technician can only post a few parts from a combination of his or her van and the consignment and other parts must be posted from the main warehouse that would indicate that FTC could not be achieved. The most important feature in this group is the parts posted from the main warehouse variable and it tells the biggest story. If few parts are posted from the main then chances are high that FTC could be achieved, while if many parts must be posted from the main then FTC is not likely.

While most of the selected features provide good insight into what drives FTC, some of the variables are not so clear. Years of experience being the top feature that does not provide much valuable information about FTC at least when looked at alone. This feature along with the number of parts postings, "is mobile WO," and "action code Service Call" features are distributed evenly around the zero line.

Some features out of the top 20 list were surprising such as the branch features that made the top 20 list. Branch 106, the highest on the list, shows a high chance for FTC even though it is not the largest or the smallest branch. However, that branch does most of the new equipment is installed in that facility, which drives up the FTC percentage. Branch 102 is also on the feature list and is very surprising, because it is in an area with many different equipment models. Also, geographically it is spread out and is plagued by high traffic, but that might be why it is a good indicator of FTC. And lastly branch 104 made the list, but negatively it seems. This seems interesting since it is not the largest or smallest branch and should not have any factors that would affect it differently than any other branch.

5 Discussion

First-Time Completion (FTC) is important to the service industry for many reasons. Customer satisfaction is one of those reasons and at the top of the list for any business providing this type of service. First-time completion drives customer satisfaction by providing reliable and quick repairs to equipment that has broken down, which makes the customer less productive. Equipment that is down or in a reduced capacity for an extended period costs the customer in many ways. Therefore, ensuring FTC on all customer repairs helps to drive up customer satisfaction, solidifies the relationship with the customer, and increases profits for the service center.

This research investigated a specific service company that provides services to warehouses and logistic centers in a specific geographic area. The research provided a list of the aspects of determining first-time completion. In doing so, the service company will be able to provide service work and achieve first-time completion on many more repairs more reliably. Considering technician experience, inventory management, the equipment, and the work to be conducted; the service center can achieve FTC on a higher percentage of work orders.

5.1 Technician Experience and Efficiency

Sending the correct technician for the repair to be completed is a key factor in achieving FTC. The number of years a technician has worked in the industry is not necessarily a good indicator of the knowledge of the equipment; the service center must also consider the technician's experience with the equipment that needs to be repaired. Even a technician with many years of experience in the field might have extremely limited knowledge of the equipment that needs repair. This research found that if a technician had a low skill level on the equipment, then FTC was less likely to be achieved even if they were highly experienced in the field. Another part of technician experience pertains to the feature that best determines FTC, which is the labor hours posted. An experienced technician both in years and on the equipment being serviced will post less labor to the work order than a less experienced technician and therefore have a higher chance of FTC. In the best performing ensemble in this study, two of the five features that showed the greatest SHAP value can be tied to the technician's experience. The SHAP value for the feature that classifies technicians as low skill level showed to be a strong predictor for not completing jobs as first-time completion. Similarly, the SHAP value for the feature that represents a technician's years of experience showed a positive relationship with first-time completion. As technicians become more experienced, they tend to complete jobs under the constraints of first-time completion more often. While this is a natural and expected correlation this still may be a call for action to improve new-hire training and continuing education for employees.

5.2 Equipment

Equipment is another factor that was found to be important in accomplishing FTC. Equipment the same brand as the servicing dealer sells and supports was found to ensure FTC more often than on competitive equipment. This is the expected result as the technicians are trained in and have more experience with the servicing dealers' brand of equipment. It is also much easier to acquire and stock parts for equipment that the service center promotes and stocks.

5.3 Inventory Management

Inventory management is usually considered to be the most impactful for generating first-time completion work. That is because a technician cannot complete the work the first time if he or she does not have the part to replace when they need it. Not only do the parts need to be in stock, but they need to be in stock in the right location and at the right time. Three features related to first-time completion ranked in the twelve highest SHAP values. The SHAP values for the number of parts a technician used from his van and the number of parts used from consignment showed a positive relationship with jobs that were completed under first-time completion. The SHAP value for the number of parts used from the main warehouse showed a negative relationship with jobs that were completed under first-time completion. While this relationship was expected, the models in this research do highlight that further research into improving inventory management can be highly valuable as there are some clear efficiencies to be gained.

5.4 Technician Write-Up

The technician write-up is the communication of the work performed by the technician that also includes the customer’s description of the issue, the actual description of the problem affecting the unit, and the troubleshooting steps the technician did to correct the problem. Using NLP to inspect the write-up it was found that some specific formatting had been implemented on each write-up and was uniform in its construction. Three keywords could be found in each write-up: Concern, Cause, Correction, therefore the NLP algorithm made use of these to develop groups to incorporate into the final models. It was suspected that the write-up would contain many insights into determining first-time completion. Out of the over 3,000 features in this study, eight of the 46 features selected through Lasso were NLP features. Three features from the NLP groups, one from the cause group, and two from the correction group were features in the 15 highest SHAP values. This was previously data that was only used if someone read or through the notes. Starting with the cause, if a work order repair was considered a cause group 0 repair, then it was more likely not to be able to be completed on the first trip. This means repairs where the technician did not specify the cause. The correction group “Pump/Motor” also showed lower first-time completion SHAP values when present. This makes sense as pumps and motors are big and expensive to stock on the van or in a consignment. And finally, the correction group “Battery” was a good feature for determining first-time completion would happen.

5.5 Implications

Completing this research allowed for a greater understanding of first-time completion, and because of that some suggestions to increase first-time completion can be made. The first suggestion would be to further structure the technician write-up, if possible, provide a select set of options to select from when determining the cause and the concern. This will not be an option for the correction; however, any standardization would help, and further NLP work would provide even better results. Another suggestion would be to create routes for each technician and identify the equipment along each route, provide extensive training on the equipment contained along the route, and stock the van and any consignments along the route with most used parts for the equipment in the route's vicinity. It is understood the routes and the technician that services that route would eventually change and be modified over time. However, keeping up with the equipment on each route is paramount to being able to achieve first-time completion.

5.6 Limitations and Challenges

Time and data were the biggest limitations and challenges. Data was a challenge because of the large volume of data made available to us. The initial dataset contained over 80 unique features and almost 3 million rows. The time to inspect, clean and process the data would have taken more than allowed for this research. Therefore, we asked the subject matter experts for a scaled-down dataset that would provide the needed criteria for determining first-time completion. In addition to the subject matter data reduction, this study also reduced the dataset through Lasso prior to modeling. With additional time this step could become redundant. Another item that was a challenge was the NLP analysis on the technician write-up, as each service center location has a unique way of completing the write-up. This provided some aspects of the analysis that were difficult to achieve. If more time were given for the research, the dataset could grow in features and rows and a unique method of NLP analysis could be developed specific to the forklift service center language.

5.7 Ethics

Ethical concerns were considered when designing the method of gathering the data and what information the data contained. Removal of all identifiable information from the dataset ensured no “data leakage” during the experiment. However, these research findings could be used to affect the compensation of technicians, team leaders, trainers, coordinators if someone chose to use the findings that way. Basing workforce promotions, compensation, or bonuses on first-time completion, is risky and ill advised, because so many factors can affect this metric. Also, these models derived from this data might not be applicable to other service centers with different technicians or equipment and therefore be unsuitable as a factor for compensation.

5.8 Future Research

This study identified the technician’s skill level and experience as key indicators of first-time completion. Further research into what technician skills is lacking and how the gap in experience can be bridged should be highly profitable both for the company sponsoring the research as well as the technicians and their career development.

The models used in the study lend themselves to predict first-time completion. Using a dataset with only data points that are available before a technician starts a job could be applied to the models in this study to create a pre-service first-time completion predictor. Such a model could be used to alert technicians to problematic jobs ahead of time and drive efficiency.

6 Conclusion

This study identified key areas that show to be strong predictors for first-time completion in the forklift industry. A wide range of service repair data, including technician write-ups and other variables, were analyzed through a variety of models with the goal of identifying the most impactful features. Identifying features allows the dealership to make informed process improvements that will increase efficiency, revenue, and customer satisfaction.

This study found three feature groups that showed the greatest importance in predicting first-time completion. The first two are the experience of the technician on the equipment, and the time it takes the technician to complete the repair. If a technician is familiar with the equipment, they are more likely to post fewer hours to a repair than another technician who is not as familiar with the same model of equipment. Since these are the top two factors in determining FTC, choosing the correct technician for the equipment that needs repair could quickly drive up the first-time completion percentage metric for a service center.

The third feature group that impacts first-time completion is equally as important as the first two are the parts carried and stocked on the service van or in the service consignment. If the technician does not have the parts needed in stock, he or she will have to travel back to a warehouse to pick up the required parts and break first-time completion. By looking at the equipment in a particular service area, the most replaced parts could be identified and stocked on the service van or consignment. Some customers might even allow for an added consignment to be stored at their location for larger or more expensive parts, that would not normally be carried on the van or consignment due to either size restrictions or fear of theft to keep first-time completion on their fleet of forklifts.

Lastly, processing the technician's write-up allowed for a better understanding of the specific repairs that contribute to first-time completion and the repairs that do not allow for first-time completion. Some repairs might not be able to be completed in a single trip, such as a bent or broken carriage, or a broken bearing in the mast of the forklift. These types of repairs can take multiple people and many hours to complete.

While some of the findings were anticipated by subject matter experts, the study highlighted areas of need for improvement that can be turned into actionable items to drive higher first-time completion rates.

References

1. Akhbardeh, F., Desell, T., & Zampieri, M. (2020). Maintnet: A collaborative open-source library for predictive maintenance language resources. arXiv preprint arXiv:2005.12443.

2. Andaleeb, S. S., & Basu, A. K. (1994). Technical complexity and consumer knowledge as moderators of service quality evaluation in the automobile service industry. *Journal of Retailing, 70*(4), 367-381. 10.1016/0022-4359(94)90005-1

3. Bratus, S., Rumshisky, A., Khrabrov, A., Magar, R., & Thompson, P. (2011). Domain-specific entity extraction from noisy, unstructured data using ontology-guided search. *International Journal on Document Analysis and Recognition, 14*(2), 201-211. 10.1007/s10032-011-0149-5

4. Corrigan, G. J., & Tsimpinos, V. (1994). *Technician PayOffs: Getting Results Through Performance Audits.* (). SAE International. 10.4271/940780 https://saemobilus.sae.org/content/940780

5. Dutta, S. (2013). *Fixing First-Time Fix: Repairing Field Service Efficiency to Enhance Customer Returns*

6. Kaasinen, E., Aromaa, S., Väätänen, A., Mäkelä, V., Hakulinen, J., Keskinen, T., Elo, J., Siltanen, S., Rauhala, V., Aaltonen, I., Hella, J., Honkamaa, P., Leppä, M., Niemelä, A., Parviainen, J., Saarinen, S., Turunen, M., Törnqvist, J., Valtonen, J., & Woodward, C. (2018). *Mobile Service Technician 4.0 – Knowledge-Sharing Solutions for Industrial Field Maintenance*. Association for Smart Learning Ecosystems and Regional Development. 10.55612/s-5002-038-001

7. Khanbhai, M., Anyadi, P., Symons, J., Flott, K., Darzi, A., & Mayer, E. (2021). Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. *BMJ Health & Care Informatics, 28*(1), e100262. 10.1136/bmjhci-2020-100262

8. Kiesmuller, G. P. (2009). A multi-item periodic replenishment policy with full truckloads. *International Journal of Production Economics, 118*(1), 275-281. 10.1016/j.ijpe.2008.08.024

9. Kiesmuller, G. P. (2010). Multi-item inventory control with full truckloads: A comparison of aggregate and individual order triggering. *European Journal of Operational Research, 200*(1), 54-62. 10.1016/j.ejor.2008.12.008

10. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

11. Nagasawa, K., Irohara, T., Matoba, Y., & Liu, S. (2014). Joint Replenishment Problem in Multi-Item Inventory Control with Carrier Capacity and Receiving Inspection Cost. *Operations and Supply Chain Management: An International Journal, 6*(3), 111-116. 10.31387/oscm0150096

12. Nee, A. Y. C., & Ong, S. K. (2013). Virtual and Augmented Reality Applications in Manufacturing. *IFAC Proceedings Volumes, 46*(9), 15-26. 10.3182/20130619-3-RU-3018.00637

13. Reichheld, F. F., & Sasser, W. E. (1990). Zero defections: quality comes to services. *Harvard Business Review, 68*(5), 105-111.

14. Salonga, G., & Aunor, V. (2021). Breathing Processing to Help Maintain Health Includes Techniques and Benefits. *Journal La Medihealtico, 2*(2), 11-15. 10.37899/journallamedihealtico. v2i2.312

15. Sinha, P. (2014). A Note on Extension of a Bernoulli Demand Inventory Model. *Operations and Supply Chain Management: An International Journal,* 54-58. 10.31387/oscm0110074

16. Matijacic, M., Fellmann, M., Özcan, D., Kammler, F., Nüttgens, M., & Thomas, O. (2013). Elicitation and consolidation of requirements for mobile technical customer services support systems-a multi-method approach.

17. Blumberg, D. F. (1994). Strategies for improving field service operations productivity and quality. *Service Industries Journal*, *14*(2), 262-277.

18. Choi, Y. H., Lee, J., & Yang, J. (2022). Development of a service parts recommendation system using clustering and classification of machine learning. *Expert Systems with Applications*, *188*, 116084.

19. Parasuraman, A., & Grewal, D. (2000). The impact of technology on the quality-value-loyalty chain: a research agenda. *Journal of the academy of marketing science*, *28*(1), 168-174.

20. Agnihothri, S., Sivasubramaniam, N., & Simmons, D. (2002). Leveraging technology to improve field service. *International Journal of Service Industry Management*.