Demand Forecasting: A Machine Learning-Based Solution to Forecast Demand Signals for Dell Services

November 2022

H19396

White Paper

Abstract

This white paper describes the features and benefits of a machine learning-based solution that predicts and estimates the future demand of various demand signals, which in turn assists several Dell teams to make better business decisions.

Dell Technologies Solutions

Copyright

Contents

Introduction 4

Methodology 6

Conclusion...16

References...16

Introduction

Executive summary

This paper explains how the Demand Forecasting Machine Learning (ML) solution helps different business teams provide forecasts several Demand Signals such as:

- New contracts and shipments forecast: Total contracts that are active during reporting fiscal week.
- Active service units (ASUs) forecast: Total unique assets that have qualifying warranties that are active during the reporting fiscal week time.
- Service requests (SRs) forecast: Cases raised or opened during the reporting fiscal week time period.
- Dispatches and tasks forecast: Parts sent based on open cases during the reporting fiscal week time period segments.

Demand Forecasting enables business teams to better understand the upcoming demands and consumption based on historical trends, patterns, and seasonality behaviors of demand and use forecast to better prepare and develop their financial plans such as Annual Operating Plan (AOP) and budget planning for the upcoming fiscal year. Also, users in operational planning roles use these forecasts to develop call center head count plans, parts procurement plans, and other plans in a more accurate and effective manner.

This paper explains additional improvement models that include classification and clustering approaches to identify the best timeseries to use based on the accuracy of the forecasts. This document also discusses how low and high volumes are handled and how the impact of COVID-19 and holidays in forecast predictions is measured.

Business challenge

Dell Technologies caters client, server, and high end server products for various countries. A good business plan is crucial for carrying out everyday business objectives. Many Dell services teams experienced difficulty preparing and developing their financial plans for the upcoming fiscal year. Operational planning for handling call center head count plans, parts procurement plans, and more also posed a challenge.

Forecasting models helps solve these problems by providing an estimation of demand signals based on historical trends, patterns, and seasonality behaviors which in provide contract and shipments count, renewals count, expiration count, ASUs count, case counts, and a dispatches counts forecast based on each fiscal week of upcoming financial year.

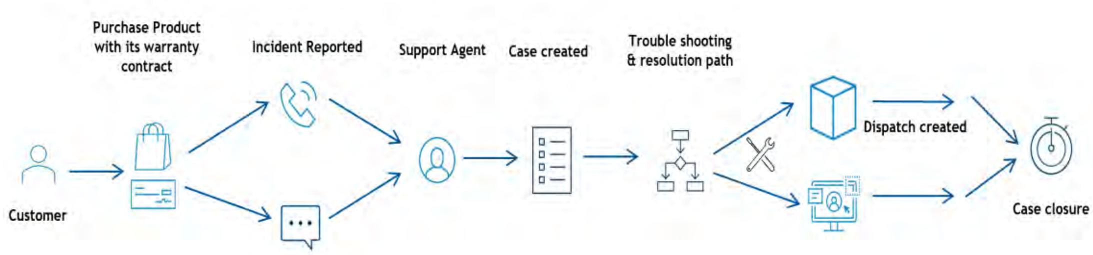

Figure 1. Contract to case creation and its closure life cycle

Document purpose

The purpose of this document is to provide an overview of the demand forecasting solution and outline the importance of employing various ML-based ARIMA timeseries models with improvement approaches to achieve efficient forecast results to help solve business challenges. This paper also uncovers ways to handle the impact of events like COVID-19 and holidays on forecast and explains how high and low volumes are handled. The evaluation of the best suited time series to create forecasts on both count and rate-based methods have also been highlighted in this paper.

Audience

This white paper is intended for business managers, data science project sponsors, analytics managers, data analysts, and data scientists.

We value your feedback

Dell Technologies and the authors of this document welcome your feedback on the solution and the solution documentation. Contact the Dell Technologies Solutions team by email.

Authors: Rohitt Punjj, Ahmed Tremo, Paras Kohli

Contributors: Will Wilson, Sarah Mostafavi, Prashant Shaw, Shilna Sasheendran

Terminology

The following table provides context for terms and acronyms that appear frequently throughout this document.

Term or acronym	Description
ASUs	Active service units are the total unique assets that have qualifying warranties that are active during the reporting fiscal week time period
APOS	After Point of Sale is the contracts made after selling different products
SRs/Cases	Service Requests/Cases raised or opened during the reporting fiscal week period
Dispatch/task /work order	Dispatches based on open cases during the reporting fiscal week time-period segments
LOB	Line of business or product category
AOP	Annual operating Plan helps team of services for budget planning for the upcoming financial year
Time Series	A sequence is taken at successive equally spaced points in time. It is a sequence of discrete-time data.

Methodology

The objective of the Dell Services teams (planning and strategy teams, finance and operation teams, and others) is to efficiently forecast demand signals. These include the following forecasts:

- **Ship**
- **Contracts**
- **Active Service Units (ASUs)**
- **Service Requests (SRs)**
- **Dispatches**

The applied data science and engineering team at Dell Technologies created the forecast of different demand singnal by employing ML algorithms that leverage ARIMA time-series demand forecasting model. Auto Regressive Integrated Moving Average (ARIMA) is a class of models that help explain a given time series based on its past values or predicts the future demand on weekly basis, considering all associated attributes such as geographic and product information, entitlements, sales channels, dispatches fields, SRs fields, and more.

To process huge historical datsets in Hadoop enviornment, Hive and Impala scripts have been leveraged and a timeseries forecasting model is scripted on R and Python. The overall process workflow is automated on SSIS packages (SQL Server Integration Services) and Apache Airflow tool and the result is published into cube for business team review.

Demand Forecasting solution components

The Demand Forecasting process consumes multiple business-specific datasets which produce the output weekly per demand signals. The streamlined workflows for these processes are described in Figure 2.

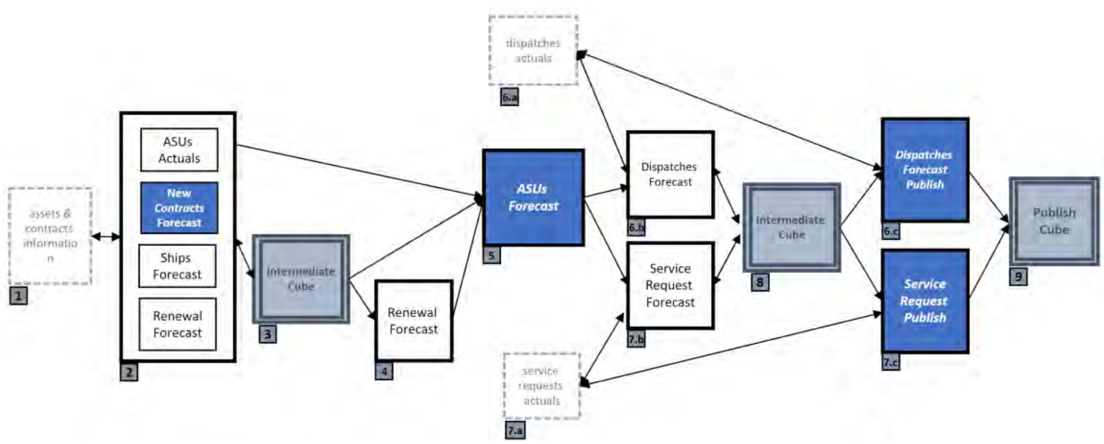

Figure 2. Workflow diagram of forecasting process

Each workflow starts with contract forecast which is later reviewed and adjusted by business teams. ASUs, dispatches, and SRs forecasts are performed sequentially. The result is moved from SQL-server table and that later published into the cube for the end users' consumption.

Forecasting model development

Model working

The codes used to run the ARIMA timeseries forecasts are available as R scripts. ARIMA is a class of models that depends on timeseries and its own past values. In this process, eight different ARIMA model variants are used to forecast forward contracts, ASUs, and more.

- **Seasonal naive** in which the last period's actuals are used as this period's forecast, without adjusting them or attempting to establish causal factors
- **Fixed ARIMA**:(0,0,1),(0,1,0)
- **Fixed ARIMA**:(0,1,2),(0,1,0)
- **Fixed ARIMA**:(1,0,0),(0,1,0)
- **Fixed ARIMA**:(1,1,1),(0,1,0)
- **Exponential smooth** a time series forecasting method for univariate data that can be extended to support data with a systematic trend or seasonal component.
- **Auto exponential smooth** (previously named forecast smoothing) is used to calculate optimal parameters of a set of smoothing functions
- **Holt linear** is forecasting method applies a triple exponential smoothing for level, trend and seasonal components

Out of these models, the one that minimizes the Root Mean Squared Error (RMSE) of the fit to a hold-out sample should be used in forecast result creation.

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with *p*, *d*, and *q* integers. In this model, integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

p: the number of lag observations in the model; also known as the lag order.

d: the number of times that the raw observations are differenced; also known as the degree of differencing.

q: the size of the moving average window; also known as the order of the moving average.

In a linear regression model, for example, the number and type of terms are included. A 0 value, which can be used as a parameter, would mean that component should not be used in the model. This way, the ARIMA model can be constructed to perform the function of an ARIMA model, or even simple AR, I, or MA models.

Autoregressive (AR(p)) Model: $$X_t = c + \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \varepsilon_t$$

Moving-Average (MA(q))Model: $$X_t = \mu + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} + \varepsilon_t$$

$\Longrightarrow$

Autoregressive, Moving-Average (ARMA(p,q)) Model: $$X_t = c + \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} + \varepsilon_t$$

In this model, *p* is the order of the Auto Regressive (AR) term. This term refers to the number of lags of X to be used as predictors and *q* is the order of the Moving Average (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model.

The SRs and dispatches are forecast forward using the same ARIMA model variants that are used to forecast contract and ASUs.

SRs and dispatches are forecast forward on both count-basis and rate-basis methods. The forecast reported for each time series is selected based on the largest rate observed in the historical data; if the largest rate observed in the historical data is less than or equal to 0.005 (0.5 percent) the rate-based method is selected, but the count-based method is used if largest attach rate observed in the historical data is greater than 0.005.

The count-based method calculates the count of the forecast, whereas rate-based method calculates the rate by combining the count with the ASUs and forecasting. The rate method determines the forecast based on count volume per ASU basis. The rates used in this forecast include mean dispatch rate for dispatches (MDR) and inbound contact rate (ICR) for SRs.

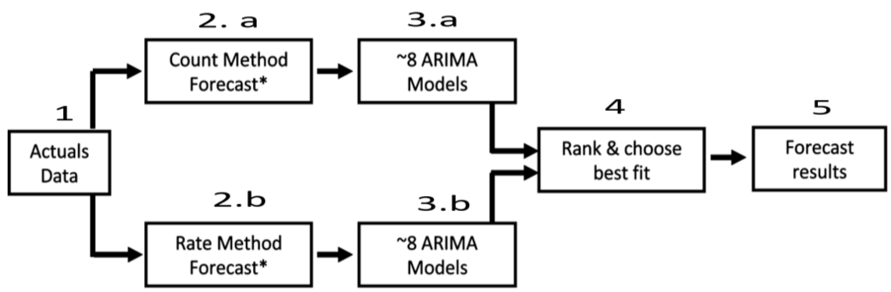

Figure 3. Flowchart of model working

Figure 4. Outcome of number of ARIMA timeseries models for dispatches counts w.r.t fiscal week

Some dispatches only occur during specific periods or once in a quarter and do not appear by statistical failures but by proactive customer engagement. Those dispatch forecasts are not generated using the ARIMA models, but it based on quarterly averages computed from the last four completed quarters.

Model improvement methods

To improve the overall computation time and incorporate forecast disturbances, such as COVID-19 in demand signal forecasting, the following model improvement approaches are being used to reduce the manual dependency of choosing the hyperparameters. This means that parameters with values that control the learning process and determine the values of model parameters that a learning algorithm.

Low volume and high volume forecast

Analysis of low and high volume forecasts showed that running various forecasting models on a low volume timeseries does not have a significant impact in the accuracy of the outcome and only increases the runtime. Therefore, it is preferred to use one simple model for the low volume and running all different models on the high-volume timeseries. This setup not only provides better results, but also saves runtime. Using low volume data reduces computation time by 80 percent and improves accuracy by about 25 percent.

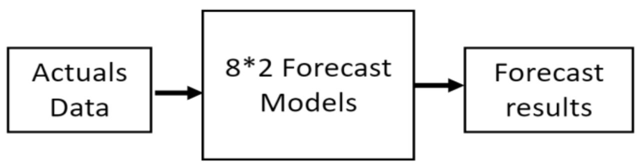

Figure 5. Flowchart without implementation of High, Low and Zero volume methods

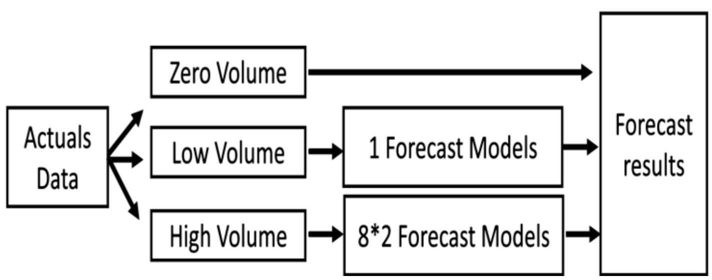

Figure 6. Flowchart after implementation of high, low, and zero volume methods

COVID-19 normalization

A COVID-19 impact analysis showed businesses required a forecast that does not incorporate variations caused by COVID-19 to accurately predict future (post-COVID-19) forecasts. A forecast before COVID-19 was used to create a normalization/delta profile for each timeseries. This forecast normalized the COVID effect before the historical data was used for forecasting. The pre-COVID-19 forecast results were used as input data for producing future forecast results. Figure 5 shows COVID-19 based data (shown in blue) and forecast results of pre-COVID-19 (shown in orange).

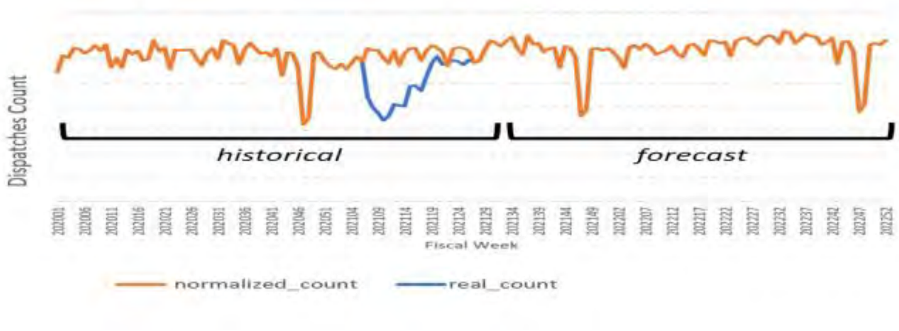

Figure 7. COVID-19 normalization for dispatches counts w.r.t fiscal week

Holiday significance

Holiday significance analyses show large variations in count volume in all demand signals such as ASUs, SRs, and dispatches. The date of certain holidays can vary each year which in turn changes the date of the demand signal variations that occur each year. Initially, timeseries are created without including holidays as a feature component. Relevant holidays provided by businesses are used as external regressors to produce the final timeseries which incorporate the effect of holidays in each timeseries.

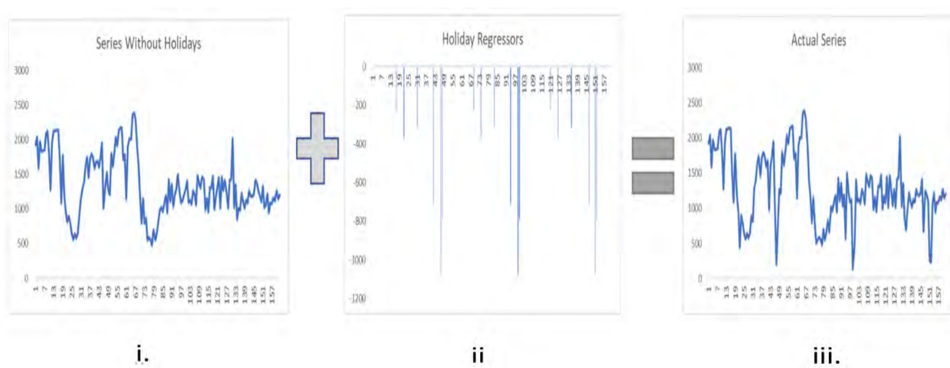

Figure 8. Representation of pre and post timeseries after including holiday regressors

Discontinuity detection and correction

Data quality and other factors negatively impact the quality of the forecast and cause significant shifts and discontinuities in historical data. Problematic timeseries are detected by dividing the timeseries into several segments and checking the variations of the counts. Drastic variations are likely caused by discontinuity in the timeseries. These discontinuities can be autocorrected by using normalization methods to feed corrected data into the forecast model.

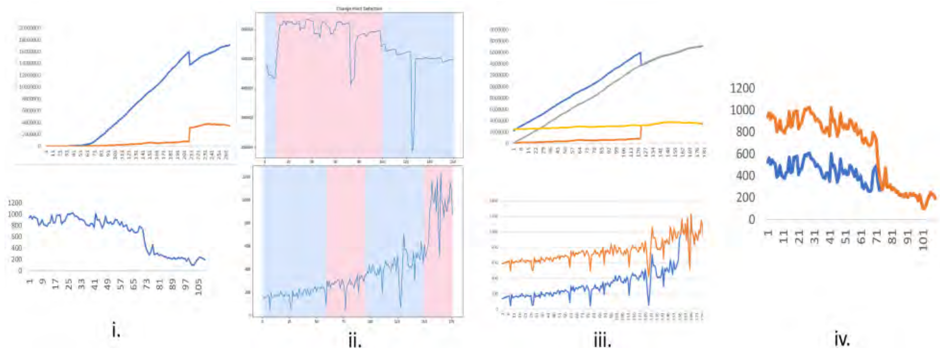

Figure 9. Representation of problematic discontinuous data and auto correcting it with cleaned data

- 9i shows discontinuity in timeseries.
- 9ii detects the point where the discrepancy occurs in the timeseries.
- 9iii shows correction or feeding clean data to make the timeseries smooth.
- 9iv shows the final output timeseries.

Dynamic threshold settings

Dynamic threshold settings are driven by an ML model that creates threshold for each timeseries dynamically based on accuracy of the method (such as rate based and count based.)

Classification and clustering techniques are used to understand which set of time series should use count-based and which set should use the rate-based model, depending on the accuracy of the forecast.

Classification using the Random Forest, LightGBM, and CatBoost classifier approaches from which CatBoost has accurate precision (60 percent) and recall (80 percent) results.

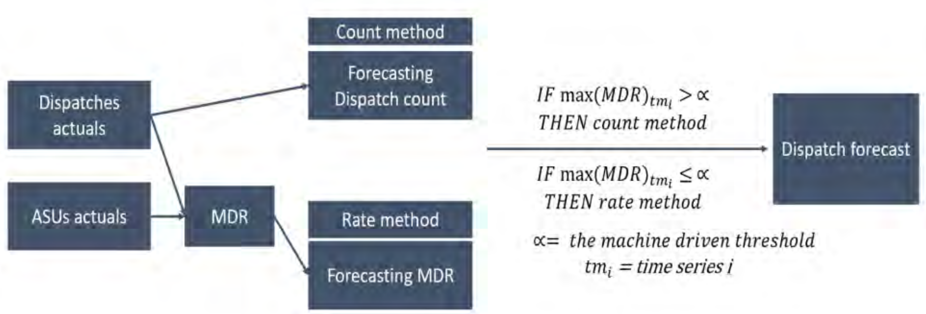

Figure 10. Process to identify the threshold value (α) which helps choose the best method

Results and discussion

The demand forecasting of all signals including ASUs, contract, renewals, dispatches, and SRs are based on each product's availability in each segment is conducted based on real data. These results are analyzed to inform future forecasts. This study examines the effectiveness of each timeseries. The final results are displayed in the following graphs.

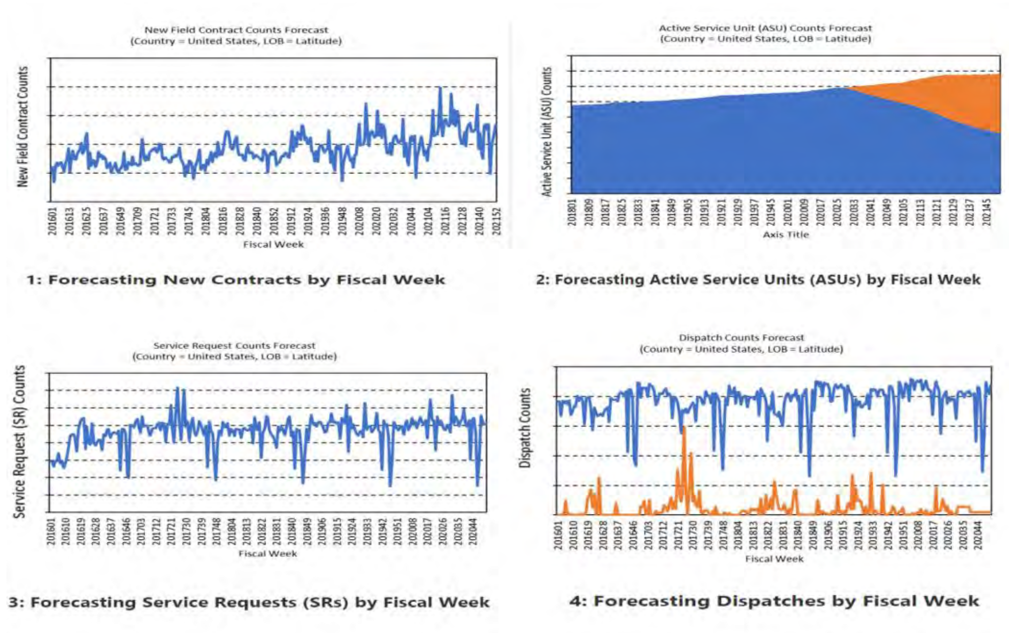

Figure 11. Final forecast results for all demand signals

We used many forecast model evaluation metrics to evaluate the model accuracy, such as:

- Mean error (ME): ME - average of all the errors in a set
- Mean absolute error (MAE): MAE - absolute values of errors
- Mean percent error (MPE): MPE - average of percentage errors by which forecasts of a model differ from actual values of the quantity being forecast
- Mean absolute percent error (MAPE): MAPE - average of the absolute percentage errors of forecasts
- Mean absolute squared error (MASE): MASE- mean absolute error of the forecast values, divided by the mean absolute error
- RMSE: the model selected is the one that minimizes the RMSE of the fit to a hold-out sample that conveys the standard deviation of the residuals, meaning it describes the concentration of the data

The final forecast reported for each LOB is the one that fits the time series the best over the last quarter.

tm_srs_id	modeling method	me	rmse	mae	mpe	mape	mase	acf1	nonzero_mape	twentyfive_perc_ape	median_ape	seventyfive_perc_ape
19284	seasonal.naive	-0.14	0.509	0.236	-9999	-9999	0.752	-0.01	36.36048	9081441.026	9081441.026	9081441.026
19284	holt.linear	-0.12	0.37	0.326	-9999	-9999	1.04	0.2	73.892899	26074286.44	26664954.11	27255621.77
19284	ses	-0.14	0.382	0.349	-9999	-9999	1.111	0.21	71.464159	28477626.87	29525485.45	30573344.03
19284	auto.exp.smooth	-0.14	0.382	0.349	-9999	-9999	1.111	0.21	71.464159	28477626.87	29525485.45	30573344.03
19284	fixed.arima:(1,1,1),(0,1,0)	-0.18	0.52	0.266	-9999	-9999	0.846	-0.01	37.640409	12815244.57	12816293.81	12816293.81
19284	fixed.arima:(1,0,0),(0,1,0)	-0.14	0.509	0.236	-9999	-9999	0.751	-0.01	36.37322	9081428.61	9081441.026	9081441.026
19284	fixed.arima:(0,1,2),(0,1,0)	-0.18	0.52	0.266	-9999	-9999	0.848	-0.01	37.618098	12854295.25	12854295.25	12854295.25
19284	fixed.arima:(0,0,1),(0,1,0)	-0.14	0.509	0.236	-9999	-9999	0.751	-0.01	36.36048	9081441.026	9081441.026	9081441.026

Figure 12. Evaluation matrix of timeseries based on single unique-id assigned to each timeseries

Key takeaways

Leveraging historical insights

The following data analyses were performed to produce efficient forecast results:

Low volume and high volume analysis

While working on different demand signals like SRs and dispatches, different volume patterns like drops and spikes w.r.t fiscal-weeks were observed. Not all timeseries have the same amount of volume and impact on overall trend. Some timeseries have consistent volume while others have sporadic demand. It is not necessary to apply the same models to every timeseries.

As shown in Figure 14, 95 percent of the volume of SRs falls in a low volume category.

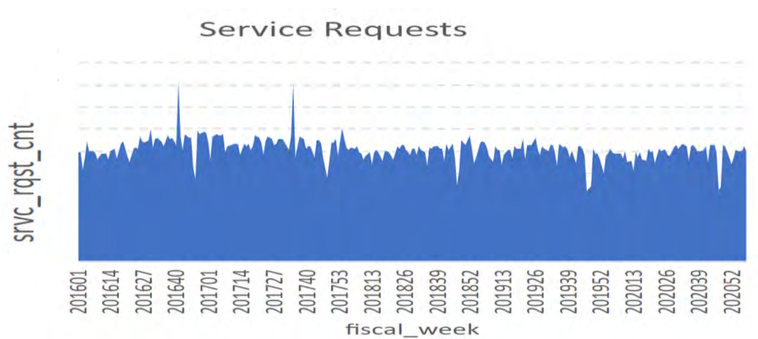

Figure 13. SR counts are presented w.r.t fiscal-week without any volume separation

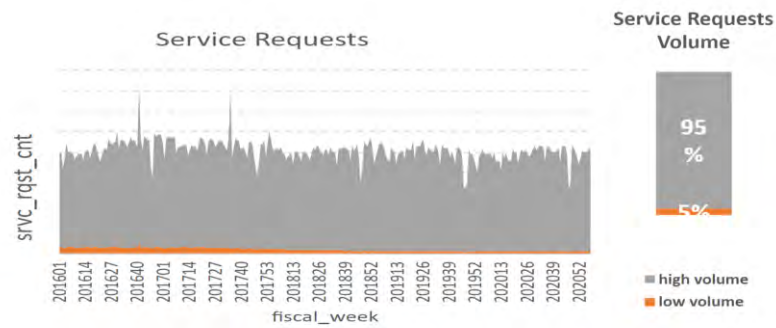

Figure 14. High and low volumes representation as percentage count distribution

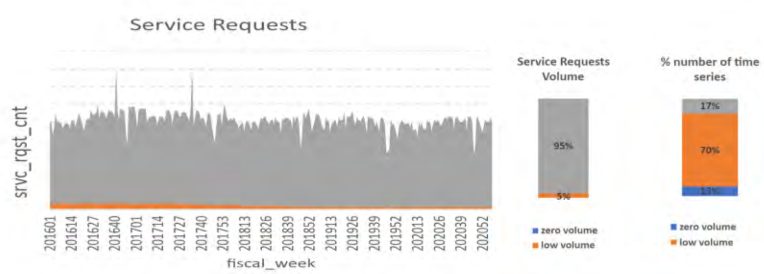

Figure 15. High, low and zero volumes representation as percentage number of time series

	SRs	Dispatches
Zero Volume	13%	12%
Low Volume	70%	70%
High Volume	17%	18%

Figure 16. Percentage distribution of SRs and Dispatches

Analysis: COVID-19 impact

COVID-19 caused a drastic disruption in supply and demand which resulted in around a 40 percent drop in dispatches and SRs that would impact the forecast results. Figure 15 shows the drop in dispatches count which has been normalized to get the accurate results for post COVID-19 forecasts.

Figure 17. Dispatches count compared to fiscal weeks

Analysis: Holiday significance analysis

During all public holidays, large variations can be seen in count volume for all demand signals like ASUs, SRs, and dispatches. Some holidays only impact certain regions. For example, Christmas is significant in US while some other countries do not experience any variation around this holiday.

Variation among lobs, which is a product that tends to have a high holiday impact, is also significant.

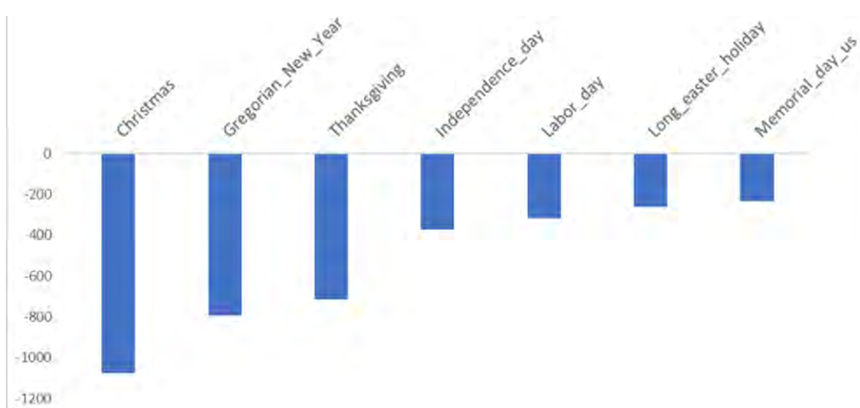

Figure 18. Holiday list vs. demand signal volume count

Impact of insights

The results of this solution show the importance of creating high and low volume-based forecasts. COVID-19 normalization analysis and holiday significance analysis helped reduce compute time by 80 percent and achieve accurate results for post-COVID-19 forecasts and capture the variation caused by holidays each year respectively.

Conclusion

Customers today expect effective products and hassle free on-time services. These expectations could not be met without a strong supply-chain that involves strategic planning that includes demand forecasting.

The solution in this white paper is a statistical and ML-based solution that creates timeseries regarding each product and its entitlements based on geographic locations. The inputs of renewal rates and holidays based on each country or region helped generate accurate results by count and rate-based forecast on weekly basis. These forecasts assist the business in parts procurements and help budget planning for each financial year.

The Demand Forecasting project was originally used by services finance and part planning teams. But it has the potential to broaden its horizon by expanding the scope of the forecasting project and changing the granularity of forecast with expanded end users.

References

Documentation

The following documentation provides additional information:

- Introduction to Time Series and Forecasting by ARIMA Model
- Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions
- Forecasting of demand using ARIMA model