

Failure Prediction Model for Predictive Maintenance

^{1st} KamalaKanta Mishra
Senior Data Scientist, IBM Corporation
Bengaluru, India
kamalm01@in.ibm.com

^{2nd} Sachin Kumar Manjhi
Senior Data Scientist, IBM Corporation
Bengaluru, India
sachin.manjhi@gmail.com

Abstract — As financial organizations strive to deliver superior omnichannel customer experiences, they are transforming their branch environments with latest digital technologies for ATMs, Branch platforms, self-service devices and other branch technologies. Simultaneously, mixing new with older, installed technologies from multiple vendors can create complex maintenance challenges. One could opt for each individual vendor's solution, but this can add complexity and may not put the crucial needs of the customer first. To maintain a customer-centric approach that leads to a high-quality brand image, improved customer satisfaction and ultimately a better bottom line, there is a need for service-oriented, vendor focused approach on delivering an integrated maintenance and technical support strategy, so that concentration on customers can be accomplished. In this direction, predictive maintenance plays a very vital role in enabling financial organizations to drive their ATM and branch business effectively to create maximum impact through predictive maintenance leveraging predictive analytics and machine learning technologies. We propose a method and Machine Learning model that takes various input data and determines likelihood of failure at a device and its component level within a stipulated future time-period with certain accuracy and precision for financial clients.

Keywords — *Predictive Analytics, Predictive Maintenance, Machine Learning, Machine Intelligence, Failure prediction model, ATM failure prediction, Regularized Random Forest (RRF).*

I. INTRODUCTION

There is no method or system available to optimally predict failure of devices and its components using machine learning by considering multiple algorithm approaches and providing the most effective approach to address the problem. Secondly, manual prediction based on historical data or trend only, leads to inaccurate result, incorrect decisions, thus impacting availability issues of a device and huge loss to firms that impacts “availability of business” and leading to customer dissatisfaction. Thirdly, IT support taskforce of any firm will have multiple levels of support and there is a tremendous need to determine optimal deployment of field engineers to tackle problems related to devices that are installed across various geographies based on effective prediction.

We propose a method and Machine Learning model that takes – a) Service Tickets (logged for a problem of a specific component of a device), b) Event information or Error codes (for a problem against a particular component in the device), c) Inventory data (Device level information regarding it's

make, model, component list, configuration parameters etc.), d) Transaction information (number of transactions, type and amount of transactions etc.), e) Weather information (temperature, humidity etc. that impacts health of a device / apparatus), f) SLA / Customer satisfaction (SLA level for a device in a region / location etc.), g) Priority usage etc. (Priority based on type of usage for business need) as input and determines likelihood of failure at a device and its component level within a stipulated future time-period with certain accuracy and precision for financial clients.

It is a “predictive engine” that looks at various parameters such as Service Tickets, Event Code / Error Codes relating to a component of the device, Inventory of the device, Installation date, Age of device, Number of Transactions till date for a particular device, Type of transactions performed, weather related parameters such as temperature, humidity etc. to provide a quantitative understanding of device health check and accurately provide when the device will break down and more importantly provides an ability to predict component of the device going to fail and underlying causes that may potentially occur in future.

II. EASE OF USE

As part of initial stages of this development, the solution is designed to predict the likelihood of failure of components of ATM devices of multiple make and model within a stipulated future time-period with certain accuracy and precision for financial clients. Then it can be replicated for any other devices for other industries.

To build a predictive model, the system is capable of extracting ATM data (ATM Inventory, Error Logs, Transactions, Service Tickets) and other external data like weather and location data. This data is then transformed to create variables(features) that are leading indicators of failure of ATM's components. The system then uses several machine learning algorithms to predict failures, compares the accuracy and other performance parameters and chooses the best algorithm for the given dataset. This trained model is then used for predicting likelihood of component failure with newer data. The model results can then be made available for downstream systems to display the results graphically for end user consumption.

III. OVERVIEW OF PROPOSED APPROACH

The proposed approach, data preparation for feature engineering process, data flow layout, proposed architecture is described below.

A. Optimal Failure Prediction of a Device

A method and approach is created to optimally predict failure of devices at a component level (such as component1, component2, component3, component4, component5, component6 etc.).

The system will consider a machine learning approach of algorithm to determine the probability of failure at every component level:

a) based on a fixed time window

b) based on dynamic time window

and will provide an optimal prediction output in terms of probability of failure from 0 to 1.

The system can be executed at a monthly level or bi-monthly level or weekly level (depending on frequency chosen) to provide predictions for a device and against each and every component / module as applicable.

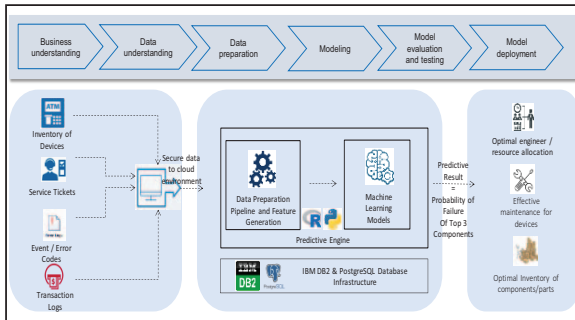
Based on these predictions, resources or field engineers will be allocated optimally to perform maintenance tasks, they can also be allocated depending on number of preventive tasks required for a region.

As part of initial versions of MVP roadmap, various parameters and features are used to come up with the model and the same is deployed in production for clients. (e.g. parameters and input data related to inventory of device/asset, service ticket information, error / event log information, transaction information).

B. Data flow layout

The input data sources involve inventory details of each device, service tickets, event / error logs and detailed information associated with each failure, transactions that occur on each device as represented below.

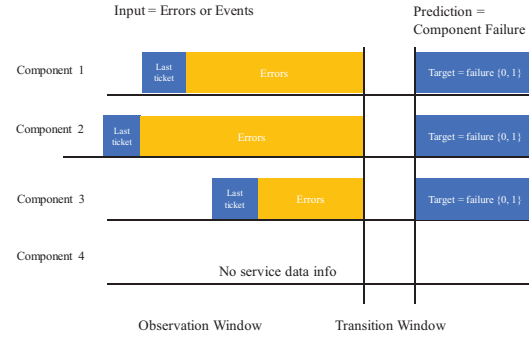
These are then used by the method with the help of feature engineering to arrive at failure predictions and recommendations at every component level.



C. Data preparation approach

The data is prepared based on fixed and dynamic time window. The approach takes errors / events as inputs and predicts for a component level failure as target variable. Binary classification is performed on prediction outcome and target variable is either 0 or 1 (either a failure occurs for a device for a component or no failure). For every device, for each component, number of errors are counted for every error code from the last ticket raised for that device and that particular component. This way datasets are stacked up month wise backwards. For fixed time window, the error count is computed considering last 30 days and for dynamic time window, it is the last ticket from which the error counts are computed.

There are three windows – observation window, transition window and predictive window which are defined to prepare features accordingly. For example, we have used predictive window of 4 weeks (around 1 month in future) for generation of prediction, transition period of 1 week to ensure when the predictions are generated every month the input data received, and actual predictions generated have a gap of few days to factor that delay. The observation window is the past historical data which is a minimum of 9 to 12 months.



D. Architecture layout

The high-level architecture layout is represented below. The microservices layers provide scalable and extensible architecture for execution of machine learning model.

Different services include different layers of architecture – one covering a main orchestration service where all configuration parameters are defined, another preparation service which encapsulates all input data into a specific json format, another processing service where all variables are features are being prepared, another predictive service where machine learning trained models are being invoked and executed.

The main orchestration service layer manages all tasks that is carried out to plan input and output flow of information across services.

IV. FEATURE ENGINEERING PROCESS

The feature engineering process involves the following:

a) Errors or event codes are used as features, b) every error id would represent a type of problem area, c) Count of errors are computed for a particular device id, for a specific time window.

Additionally, other features such as age of device (based on purchase date, installation date), make / model of device, no of transactions, count of transactions etc. are also included in entire feature set in addition to these Error / Event features.

There are around 380 features which we have used out of which 360 are Error/Event id related features.

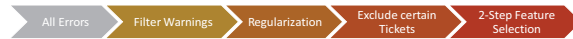
Below is an illustrative view of only error id related features that are prepared as part of the feature engineering process.

Device ID	ERR_0001	ERR_0002	ERR_0003	ERR_0004	ERR_0005	ERR_0006	ERR_0007	ERR_0008	ERR_0009	ERR_0010
Device0001	407	491	476	91	942	319	999	197	670	496
Device0002	944	58	243	451	362	720	139	931	509	263
Device0003	922	702	797	230	606	366	947	517	121	391
Device0004	861	126	977	168	669	841	506	513	409	180
Device0005	313	762	907	75	4	925	549	887	109	376
Device0006	53	655	51	47	652	902	697	773	698	32
Device0007	352	223	375	367	778	946	43	246	250	131
Device0008	675	372	445	262	751	708	295	870	293	880
Device0009	906	349	575	495	685	739	711	889	577	897
Device0010	515	615	692	237	281	297	838	719	862	272
Device0011	887	890	257	181	88	806	659	728	180	792
Device0012	49	660	597	51	693	443	845	864	401	64
Device0013	700	181	389	912	333	297	39	787	449	1
Device0014	644	918	340	910	625	810	880	262	688	518
Device0015	778	502	623	362	767	844	916	278	273	86
Device0016	962	160	558	679	254	9	796	973	624	737
Device0017	770	98	865	226	219	552	886	521	330	42
Device0018	408	850	532	245	476	662	780	668	930	52

Other factors will also be considered as part of future work – some of the features can be environmental factors, seasonality, locality, importance, usage priority etc.

V. RESULTS

The results evolved as our journey progressed based on multiple experiments. Experiments in machine learning modelling phase involved stepwise evolution of approach (but not limited to the following):



- All Errors / Events considered as feature set
- Errors / Events that are categorized as Warnings are removed from the feature set
- Regularization approach applied. RRF model used
- Less frequent tickets are removed (whose count ≤ 50), this is to ensure less frequent service tickets are ignored as they do not impact much on overall prediction outcome
- 1st step: involved all feature set, 2nd step: involved top 30~35 features based on variable importance derived from RRF

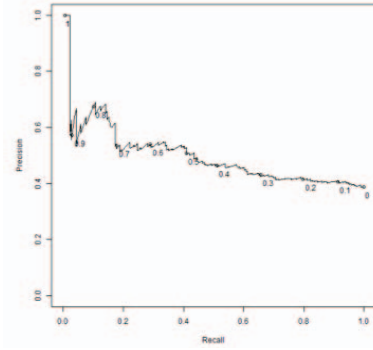
The model performance is evaluated by precision and recall values rather than accuracy.

Precision will tell what portion of time we will be correct about our prediction that device x is going to fail during the predictive window. Recall indicates that, out of all devices that are going to fail during the predictive round, what is the portion we will correctly identify. Accuracy, on the other hand, informs what portion of time we will be correct about the real outcomes.

The confusion matrix around prediction or no prediction will be as follows:

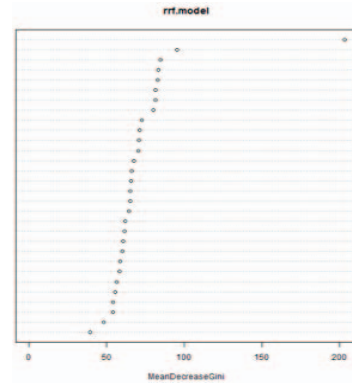
	Not failed	Failed
No prediction	True Negative	False Negative
Prediction	False Positive	True Positive

Model Output:



The precision and recall chart above is derived after several steps of experiments that are performed and mentioned above to refine the performance of the model.

The feature importance graph is derived from the regularized random forest model (for top 30~35 features) based on the Gini coefficient. The chart is shown below for reference.



Final predictive output that is generated in an excel format includes device id, probability of failure, priority and top 3 components in order of failure likelihood.

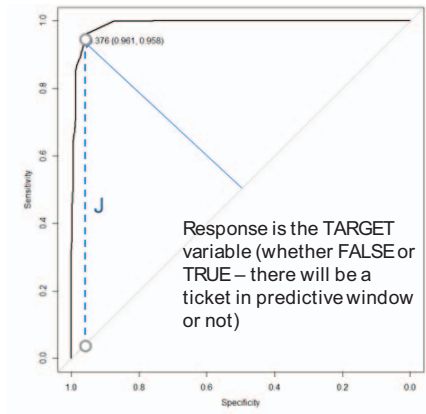
Device ID	Likelihood of a Failure		Component that may Fail		
	Probability of failure	Priority	Dev1	Dev2	Dev3
Device0001	0.792	High	Component 1	Component 3	Component 4
Device0002	0.746	High	Component 1	Component 3	Component 2
Device0003	0.744	High	Component 1	Component 3	Component 2
Device0004	0.734	High	Component 1	Component 2	Component 4
Device0005	0.688	High	Component 2	Component 4	Component 3
Device0006	0.686	Medium	Component 3	Component 2	Component 4
Device0007	0.652	Medium	Component 3	Component 1	Component 2
Device0008	0.598	Medium	Component 3	Component 1	Component 2
Device0009	0.422	Medium	Component 4	Component 1	Component 2
Device0010	0.385	Medium	Component 2	Component 1	Component 3
Device0011	0.376	Low	Component 4	Component 1	Component 2
Device0012	0.365	Low	Component 4	Component 1	Component 2
Device0013	0.338	Low	Component 3	Component 1	Component 2
Device0014	0.334	Low	Component 2	Component 1	Component 4
Device0015	0.259	Low	Component 1	Component 4	Component 3

Cut-off probability is computed based on Youden J statistic parameter. The definition of Youden J index is defined below which uses the sensitivity and specificity values to compute the cut-off.

$$J = \text{sensitivity} + \text{specificity} - 1$$

$$J = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} + \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} - 1$$

Based on the cut-off value of a device, predictive outputs are filtered accordingly. Example of the chart capturing Youden J statistic is captured below.



VI. CONCLUSION

The model is used for financial clients and has a strong proven success story already running in production. The business objective for the financial giant is to move up the value chain leveraging predictive to showcase impact in terms of increase in availability of devices, reducing downtime of devices, ability to decrease service tickets and become leading brand in their geography based on customer satisfaction. The business objectives are achieved with this and goal is to further research and scale this beyond financial clients and other industries.

VII. FUTURE WORK

As part of future work, this is expected to be experimented with additional feature sets. This will also be extended to other industries such as Retail, Manufacturing, Healthcare where device level failures can be predicted in advance with data driven outcome.

VIII. RELATED WORK AND REFERENCES

- [1] Tong Liu, P. Bahl, I. Chlamtac on - Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks [<https://ieeexplore.ieee.org/abstract/document/709453/>]
- [2] Jie Zhao, Limei Xu, Lin Liu on - Equipment Fault Forecasting Based on ARMA Model [<https://ieeexplore.ieee.org/abstract/document/4304129/>]
- [3] David Harris, Jerry Sychra, Lisa Schmit, W.Kuhrman on - Method for predicting machine or process faults and automated system for implementing same [<https://patents.google.com/patent/US20020091972A1/en>]
- [4] A.Lahyani, P.Venet, G.Grellet, P.J.Viverge on - Failure prediction of electrolytic capacitors during operation of a switchmode power supply [<https://ieeexplore.ieee.org/abstract/document/728347/>]