# ASSIGNMENT 12
# MODEL MONITORING DASHBOARD PROPOSAL

5420 Anomaly Detection, Fall 2020

- Harsh Dhanuka, hd2457

# Agenda

1. **Learn about Model Decay**

2. **Build a model performance monitoring dashboard plan, including the following aspects:**

   - Model

   - System

   - Service

   - Production Cost

**3. Explain the Dashboard elements**

Dataset: **Mortgage Loan Default**

# Problem at Hand

**MODEL DECAY** - **A model is likely to lose its predictability over time**
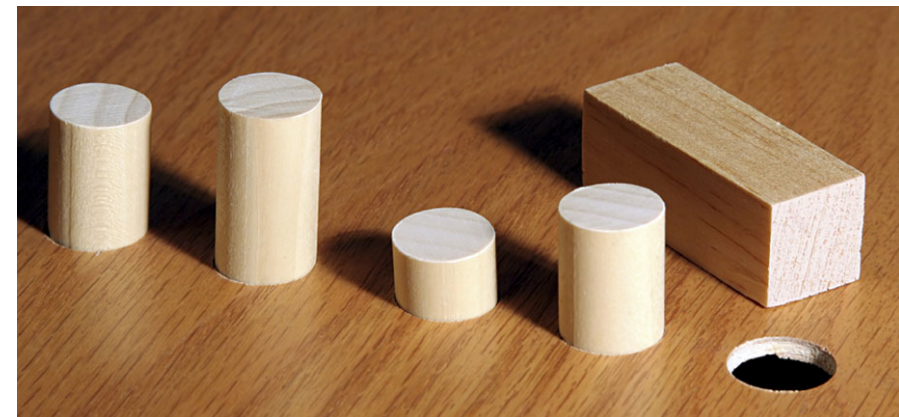
## Changes in Business Environment



## Models can be outdated



## New uses for predictions



## Data Interruption

# Monitoring Dashboard

**Analytical Solution** – Build a **MONITORING DASHBOARD** to track usage by Customers or End Users

Set threshold(s) and parameters to accept/reject the model and decide its validity and stability. Alerts that can detect failure or try to predict future failure ahead of time.

## Need to Monitor Performance

- Ensure models are "good to buy"
- Models should continue to deliver their promised outcomes
- Model should continue to detect likely loan defaults timely without failure.
- Tracking the performance of a model warrants its service to the users

## Components of the Dashboard

**Model Performance**
- Composition of Input Variables
- Predicted Values Accuracy
- Volume of Input/Output Data
- Confusion Matrix values

**System Usage**
- System Performance (CPU, RAM)

**Production Cost**
- Production Costs (Labor, Server)

**Service Response**
- Service Level Agreements (SLA's)
- Response Time, User Rating

# Variables

**Dashboard will be made up of a model and its results in real-time. For this purpose, I will be using the top 10 most important features:**

**TD013** – Count of Queries in Past 6 months

**AP003** – Education Code

**AP004** – Loan Term

**MB007** – Type of device/OS

**TD009** – Count of Queries in Past 3 months

**TD005** – Count of Queries in Past 1 month

**CR015** – Months of Credit Card Max

**TD014** – Count of Queries in Past 6 months for Small Loans

**MB005** – Number of Years phone is Active

**CD123** – Count of Distinct Outbound Calls in Past 3 months

# Proposed Dashboard

**1** **Model Performance Metrics**

This is a core part of the entire Monitoring Dashboard, which includes all the **model related aspects**. There are numerous factors to monitor in regards to the model performance:

**Metrics**

**KPI's**

- LIFT Score
- AUC Score
- PR Score
- rmse
- $R^2$
- F Score
- Number of anomalies

**Dependent Variables**

- New variables
- New levels in variables
- Composition of Each Input Variable
- Check the averages, measure drift
- Check statistics, maximum, minimum, average, etc
- Percent of missing values

**Target Variable**

- The difference between the actual and predicted variables
- Prediction values over time, their statistics such as average, minimum, maximum, etc.
- Volume of Predictions

# Proposed Dashboard

## Model Performance Dashboard Cluster

Option for :
**YTD, MTD, WTD, Daily**

| Model Performance Metrics | YTD ▼ |

**Increase vs Decrease & % of change**

Any changes outside the threshold need to be immediately reported
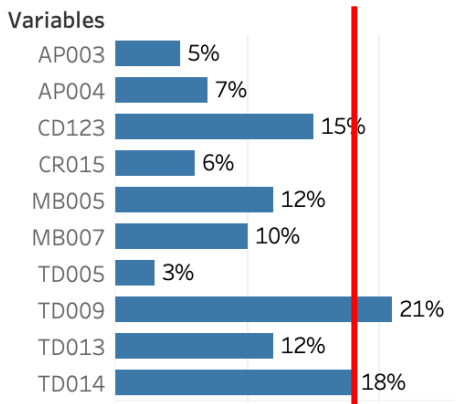
### Evaluation KPI's

**RMSE:** 0.27 ▼

**LIFT:** 2.89 ▲

**AUC:** 0.82 ▲

**PR:** 0.64 ▲

### % Drift from Avg: Top Variables

Variables
- AP003 — 5%
- AP004 — 7%
- CD123 — 15%
- CR015 — 6%
- MB005 — 12%
- MB007 — 10%
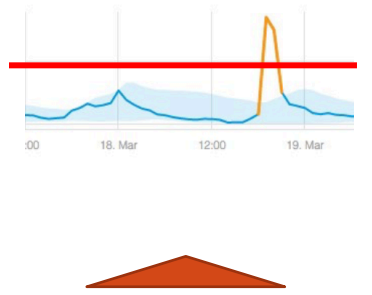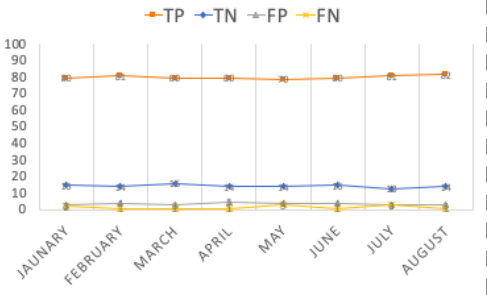- TD005 — 3%
- TD009 — 21%
- TD013 — 12%
- TD014 — 18%

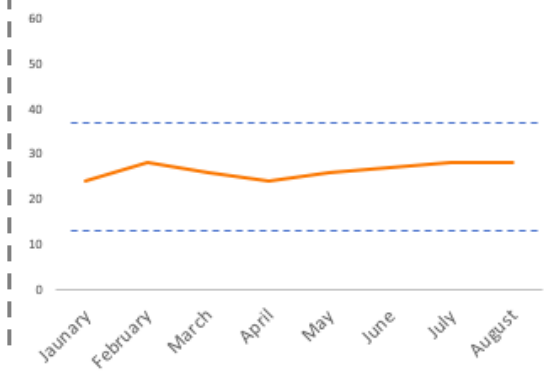**Changes in statistics - Average, etc.**

Any big deviations must be immediately reported

**Confusion Matrix**

True Positive should be highest



-TP -TN -FP -FN



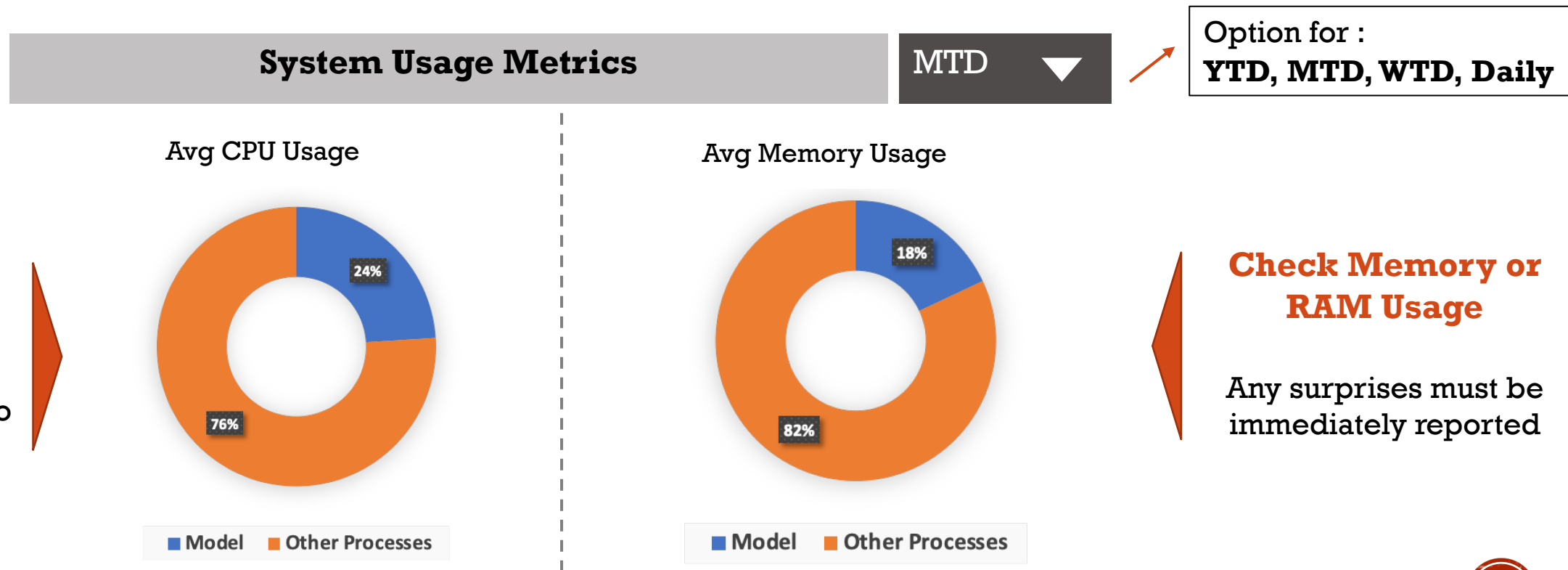**Percentage of Missing Values**



**Average Predicted Value**

Any change outside the threshold should be immediately reported

# Proposed Dashboard

**2** **System Usage Metrics**

These metrics are about the system and network related aspects. These has to do with **CPU Usage**, **Memory Usage**, **Disk**, **Network I/O**, and others. This will be monitored by the IT and Networking team.



| System Usage Metrics | MTD ▼ |
|---|---|

Option for :
**YTD, MTD, WTD, Daily**

**Avg CPU Usage**

**Avg Memory Usage**

**Check CPU Usage**

Any changes outside the set threshold need to be reported

24%

76%

■ Model ■ Other Processes

18%

82%

■ Model ■ Other Processes

**Check Memory or RAM Usage**

Any surprises must be immediately reported

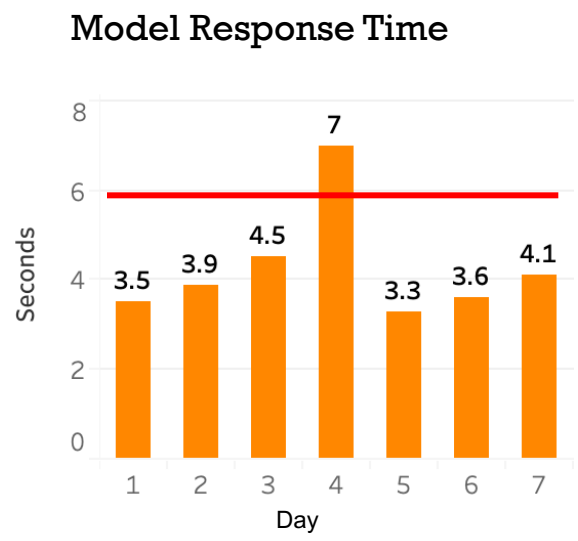# Proposed Dashboard

**③ Service Response Metrics**

Production models almost always have service level agreements (*a commitment between a service provider and a client*) that have to do with how **quickly they must produce results** and **how often they are allowed to fail**. Further, they also evaluate the **user rating scores**.



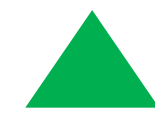| Service Response Metrics | WTD ▼ |

Option for :
**YTD, MTD, WTD, Daily**

**Model Response Time**

**Avg User Rating (1-5)**

**Check Model Response Time**

Any changes outside the set threshold need to be reported

**4.21** ▲

**Check Average User Rating**

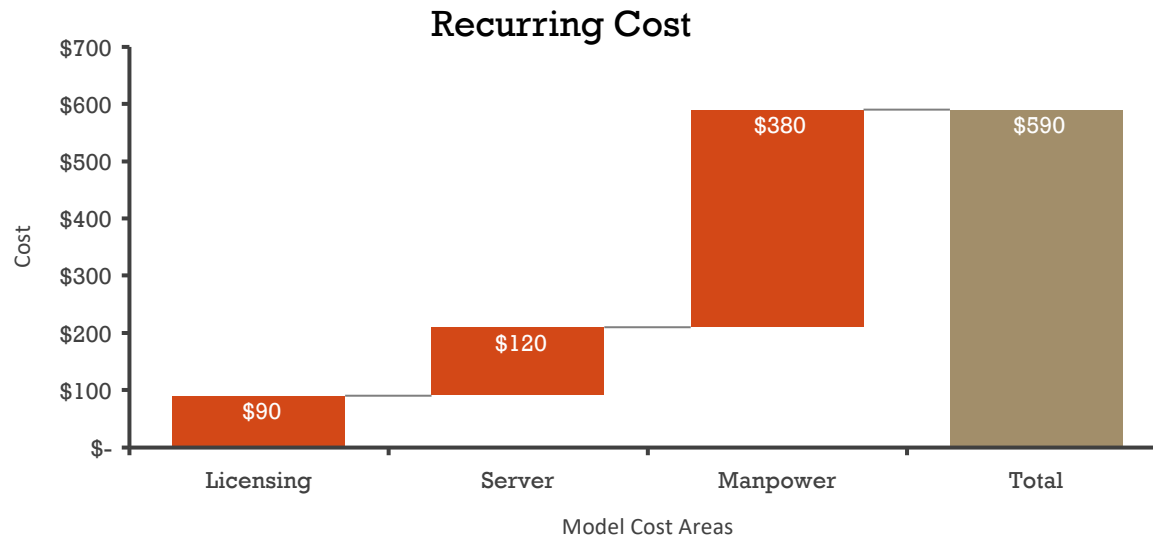Any surprises must be immediately reported

# Proposed Dashboard

**4** **Production Cost Metrics**

This relates to all the billing and cost information such as **HR cost, manpower cost, licensing cost, server and network costs**, and others.

**Production Cost Metrics**   | MTD ▼

Option for :
**YTD, MTD, WTD, Daily**

### Recurring Cost



**Check Averages of Contribution to Total Cost**

Any surprises beyond a set threshold must be immediately reported

# Loan Default Model: Performance Monitoring Dashboard

## Model Performance Metrics [ MTD ▼ ]
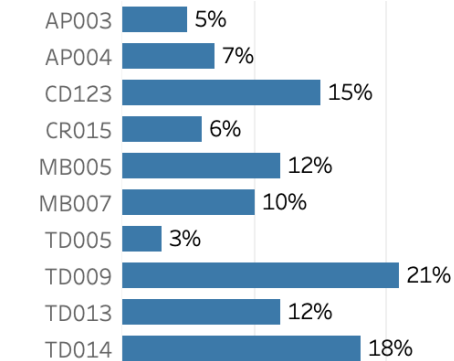
### Evaluation KPI's

**RMSE:** 0.27 ▼

**LIFT:** 2.89 ▲

**AUC:** 0.82 ▲

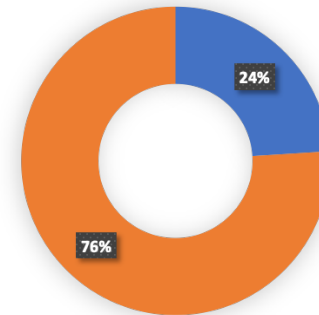**PR:** 0.64 ▲
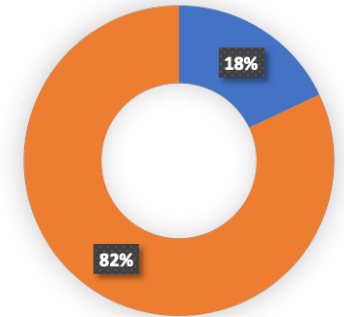
### % Drift from Avg: Top Variables

Variables

| Variable | % |
|----------|-----|
| AP003 | 5% |
| AP004 | 7% |
| CD123 | 15% |
| CR015 | 6% |
| MB005 | 12% |
| MB007 | 10% |
| TD005 | 3% |
| TD009 | 21% |
| TD013 | 12% |
| TD014 | 18% |

## System Usage Metrics [ MTD ▼ ]

### Avg CPU Usage

24% / 76%

■ Model  ■ Other Processes

### Avg Memory Usage

18% / 82%

■ Model  ■ Other Processes

## Service Response Metrics [ WTD ▼ ]

### Model Response Time

| Day | Seconds |
|-----|---------|
| 1 | 3.5 |
| 2 | 3.9 |
| 3 | 4.5 |
| 4 | 7 |
| 5 | 3.3 |
| 6 | 3.6 |
| 7 | 4.1 |

Day
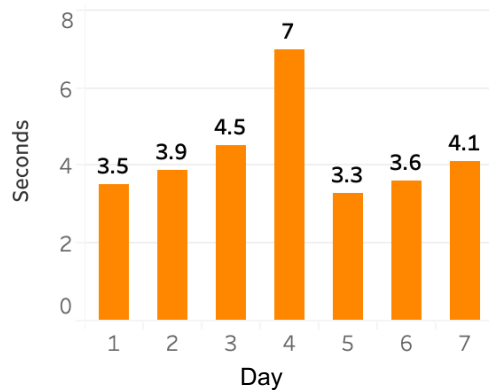
### Avg User Rating (1-5)

## 4.21 ▲

## Production Cost Metrics [ MTD ▼ ]

### Recurring Cost

Cost

| Model Cost Areas | Cost |
|------------------|------|
| Licensing | $90 |
| Server | $120 |
| Manpower | $380 |
| Total | $590 |

Model Cost Areas
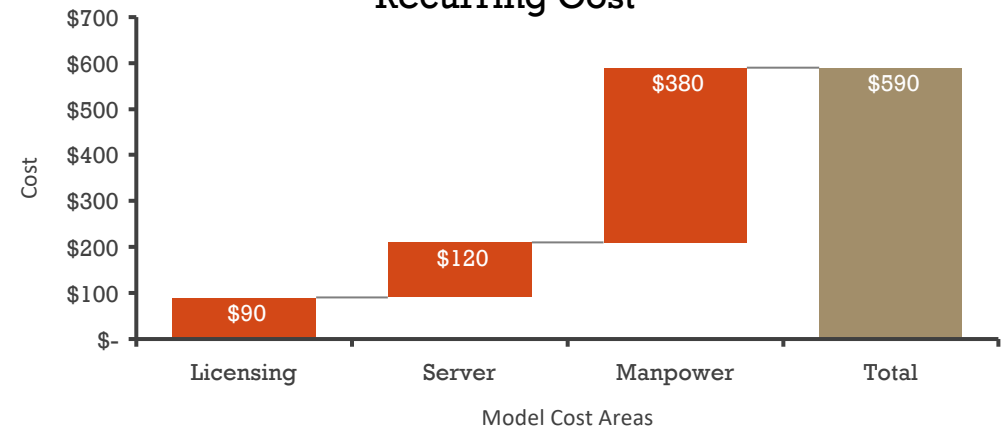
## Next Steps

The Monitoring Dashboard will give us all the necessary insights with regards to model decay, and the performance of our model over time. Maybe the business environment changed, or there are new variables, or there are new uses for the model results. **So, once we have the data in hand, what do we do?**

It is the time to consider to **update the model**. How do we do it?

1. **Re-train:** When we re-train a model, we will go back to the drawing board and follow the model building and evaluation. This is the ground-up opportunity to reinvent new use cases.

2. **Re-fresh:** This is when we do not need to go back to the first step. The model has been in service successfully with good customer ratings. All it needs to do is to add one or two variables and to refresh the parameters to calibrate to the best predictability.

# Conclusion

The power of machine learning to extrapolate ideas, predict answers, and solve problems cannot be underestimated. However, it **requires constant monitoring of our machine learning model to achieve the most accurate results.** The most efficient and scalable way to do this is by setting machines to watch our machines. Once one set of machines learns the patterns of the machines that they're observing, they can understand when their output changes in anomalous ways.

**We can't set and forget our machine learning modules, and if we do, we are likely to miss anomalies that result in lost revenue, reputational damage or worse**

# FUTURE CONSIDERATIONS

**Tools to leverage for building dashboards for Monitoring Machine Learning Models**

# FUTURE CONSIDERATIONS

**Use third party apps for built-in dashboards for Monitoring Machine Learning Models**