LEAD TIME PREDICTION USING ML ALGORITHMS

A case study by a *semiconductor manufacturer*

Presented by :
Puriwat Sangrawee / Student ID: 65056071

MAIN SOURCE :

Lead time prediction Using Machine Learning Algorithms : A case Study by a Semiconductor Manufacturer

- Written by : Lukas Lingitz et. al. (Fraunhofer Austria Research GmBH)
- published on ScienceDirect (2018)

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 72 (2018) 1051–1056

www.elsevier.com/locate/procedia

51st CIRP Conference on Manufacturing Systems

Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer

Lukas Lingitz[a], Viola Gallina[a,*], Fazel Ansari[a,b], Dávid Gyulai[c,d], András Pfeiffer[c], Wilfried Sihn[a,b], László Monostori[c,d]

[a]Fraunhofer Austria Research GmbH, Theresianumgasse 27, A-1040 Wien, Austria
[b]Vienna University of Technology, Institute of Management Science, Theresianumgasse 27, A-1040 Wien, Austria
[c]Centre of Excellence in Production Informatics and Control (EPIC), Institute for Computer Science and Control (SZTAKI), Hungarian Academy of Sciences (MTA), Budapest, Hungary, Kende str. 13-17, H-1111 Budapest, Hungary
[d]Department of Manufacturing Science and Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111 Budapest, Hungary

* Corresponding author. Tel.: +43-676-888-616-46; *E-mail address:* viola.gallina@fraunhofer.at

Abstract

The accurate prediction of manufacturing lead times (LT) significantly influences the quality and efficiency of production planning and scheduling (PPS). Traditional planning and control methods mostly calculate average lead times, derived from historical data. This often results in the deficiency of PPS, as production planners cannot consider the variability of LT, affected by multiple criteria in today's complex manufacturing environment. In case of semiconductor manufacturing, sophisticated LT prediction methods are needed, due to complex operations, mass production, multiple routings and demands to high process resource efficiency. To overcome these challenges, supervised machine learning (ML) approaches can be employed for LT prediction, relying on historical production data obtained from manufacturing execution systems (MES). The paper examines the use of state-of-the-art regression algorithms and their effect on increasing accuracy of LT prediction. Through a real industrial case study, a multi-criteria comparison of the methods is provided, and conclusions are drawn about the selection of features and applicability of the methods in the semiconductor industry.

Peer-review under responsibility of the scientific committee of the 51st CIRP Conference on Manufacturing Systems.

Keywords: Lead time; prediction; machine learning; regression methods; comparison; features

TODAY'S AGENDA

PART 1
Introduction

PART 2
Literature Review

PART 3
Method Explanation

PART 4
Result Explanation

PART 5
Future Research / Recap / Q&A

TODAY'S AGENDA

PART 1
Introduction

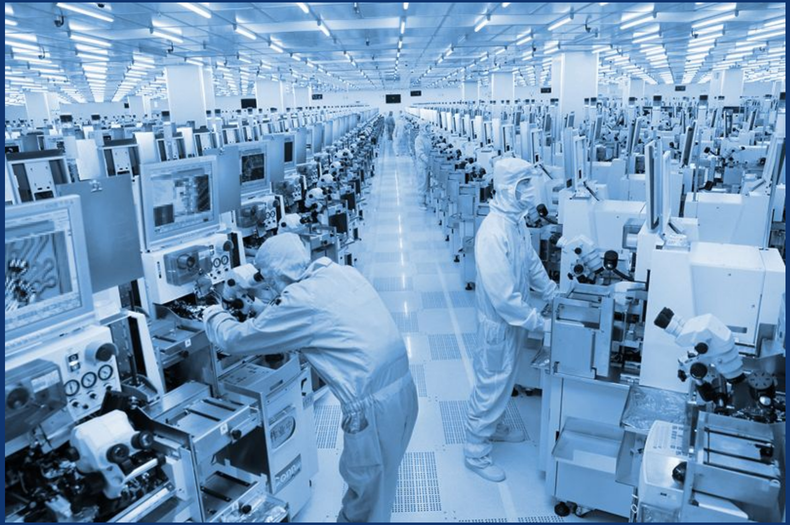
PART 2
Literature Review

PART 3
Method Explanation

PART 4
Result Explanation

PART 5
Future Research / Recap / Q&A

1.1 WHAT & WHY THEY DO ?



This paper **examine** ***Regression Algorithm*** **and their effect on increasing Accuracy of Lead Time Prediction through an real industrial case study**

Best Model & its parameter

Better Quality & Efficiency for PD Planning & Scheduling

1.2 WHAT IS *LEAD TIME (LT)* ?

Manufacturing Lead time is the time it takes to produce a product or service. It included the time duration from the start of the manufacturing process to the point where the product or service is ready for delivery

This time can be affected by a variety of factors , including :

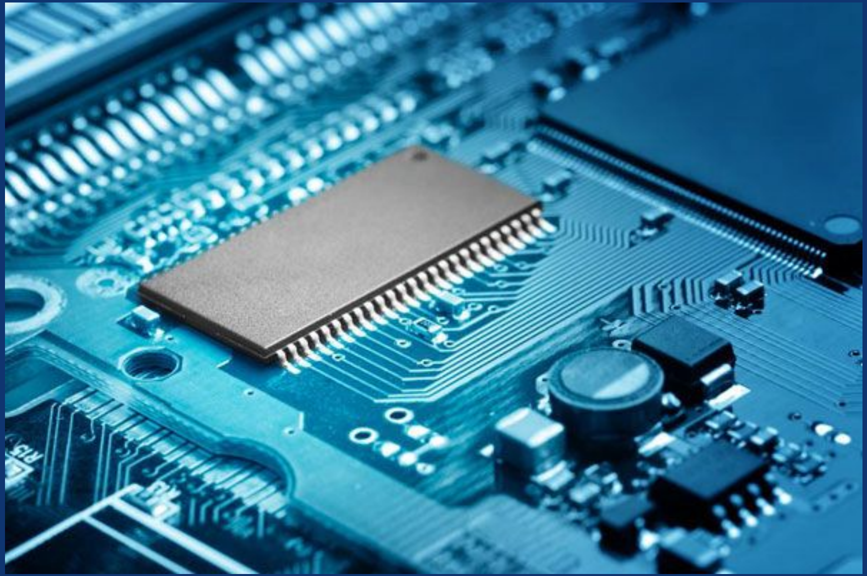
- Complexity of the product or service
- Availability of raw materials
- Number of production steps involved
- Availability of labor and machinery and
- Size of the order

1.1 WHAT IS *SEMICONDUCTOR* ?

Semiconductor are the materials which have a conductivity between conductor (generally metals) and non-conductors (insulator) (such as ceramics) .

Gallium Arsenide , Germanium and Silicon are some of the most commonly used semiconductors. Silicon is used in electronic circuit fabrication and gallium arsenide is used in solar cell , laser diodes etc.

In case of Semiconductor Manufacturing , sophisticated LT prediction method are needed due to complex operations , mass production , multiple routings and demands to high process resource efficiency



TODAY'S AGENDA

PART 1
Introduction

PART 2
Literature Review

PART 3
Method Explanation

PART 4
Result Explanation

PART 5
Future Research / Recap / Q&A

2.1 KNOWLEDGE DISCOVERY IN PRODUCTION PLANNING & CONTROL

In the past few years there have been a significant growth in the number of papers discussing the usage of different data analytics methods in production management

According to study from *Rainer et. al.* after applying different data mining techniques for converting ***big data*** to ***Smart data*** , companies have experience payback of at least 10 times of their investment

Cheng et. al. explain from their survey that the four most reflected **typical application are advanced planning & Scheduling , quality improvement , fault analysis and defect analysis** and identified PPC as a research gap in 2009 and the review in 2017 has revealed just a few application in this passed 9 years. Consequently , more attention from the research community would be needed to data mining in PPC

2.2 LEAD TIME PREDICTION WITH REGRESSION : STATE-OF-THE-ART METHODS

The literature survey revealed that the most research of **time related data mining analysis** (e.g. Flow time , (lot) cycle time , lead time)

- Have focused on the whole process flow
- Have used a dataset generated by simulation
- Have applied and compared just a few ML algorithms

Here are some example of relevant research topic…..

Pfeiffer et. al. compared result from 3 ML models predicting LT with 8 features and data gained by discredited-event simulation → Random Forest (RF) outperformed linear regression & Regression Tree model

Ozturk et. al. applied Regression Tree with simulated data source of 4 shop types in order to determine the most relevant attribute having relatively high predictive power

De Cos Juez et. al. analyzed the results of a SVM models with 8 features (reduced from 12) to predict whether a batch is going to be completed in the forecast time or not

TODAY'S AGENDA

PART 1

Introduction

PART 2

Literature Review

PART 3

Method Explanation

PART 4

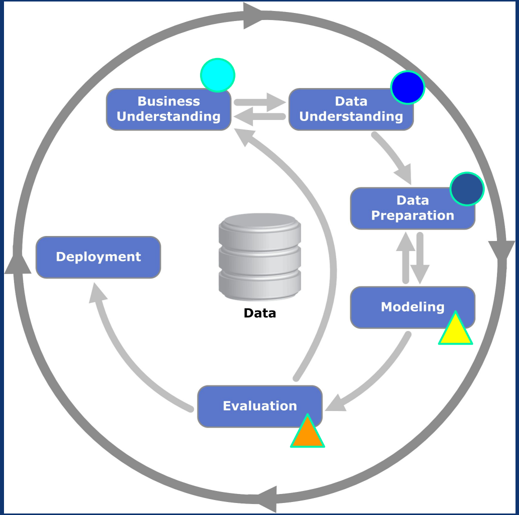
Result Explanation

PART 5

Future Research / Recap / Q&A

3.1 OVERALL PROCESS MODEL

This study reformatted original CRISP-DM to be new process step as shown in Fig 3.1



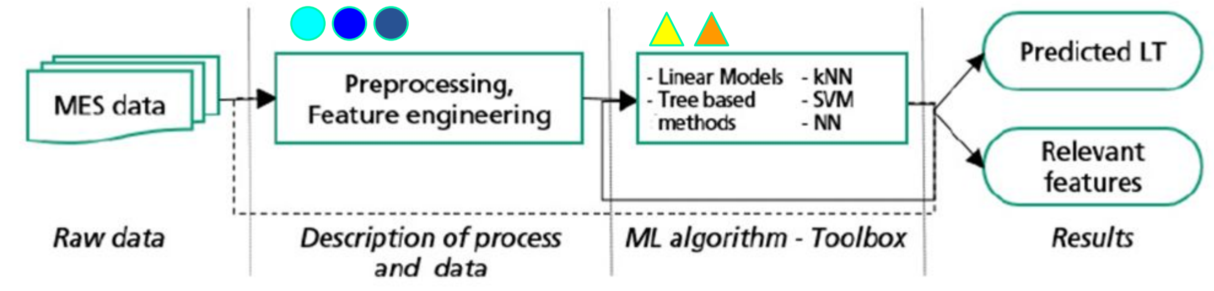
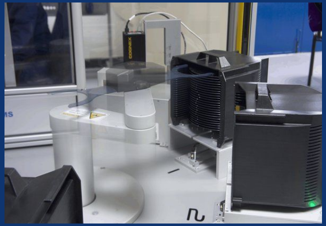


Fig 3.1 Data Analysis Process Steps applied in PPC

3.2 DESCRIPTION OF MANUFACTURING PROCESS

Lot Wafers



Sort

Bakefuse

(Buffer Zone)

Sputter

- **Sputter & Bakefuse** **process are related with** ***Maximum Laytime***
  (buffer size are limited → lot before are blocked)
- **Lead time (LT)** **should be calculated from** **both per lot & per layer**
- **a lot can arrive several time**
- **That's why we need precise Lead Time (LT) Prediction Model**

3.2 DESCRIPTION OF MANUFACTURING PROCESS

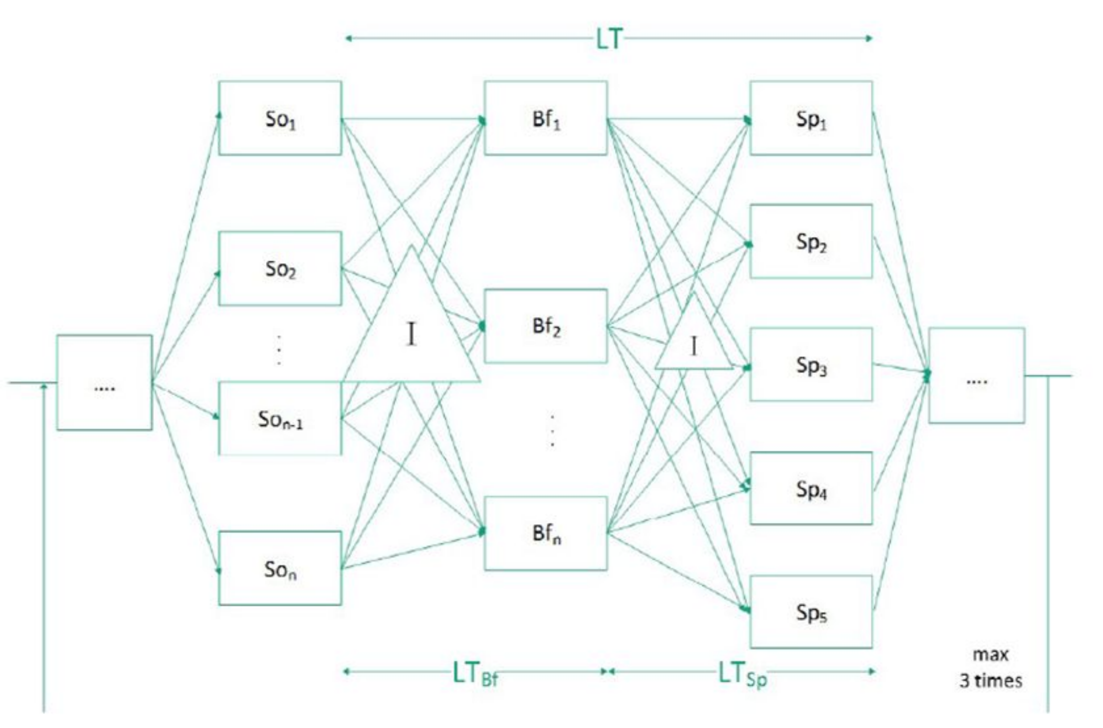


Fig 3.2 Illustration of the analyzed manufacturing process step

3.3 DESCRIPTION OF DATA

Table 3.3 Description of raw data

Name	Type	Granularity
Product number	Alphanumeric string	106
Customer	Alphanumeric string	33
Production lot	Alphanumeric string	23819
Operations	Alphanumeric string	14
Routings	Alphanumeric string	38
Time stamp	Date and time	Seconds
Production quantity	Integer	0-25
Equipment	Alphanumeric string	43
Priority	Integer	3
Status of operations	Alphanumeric string	22

3.3 DESCRIPTION OF DATA

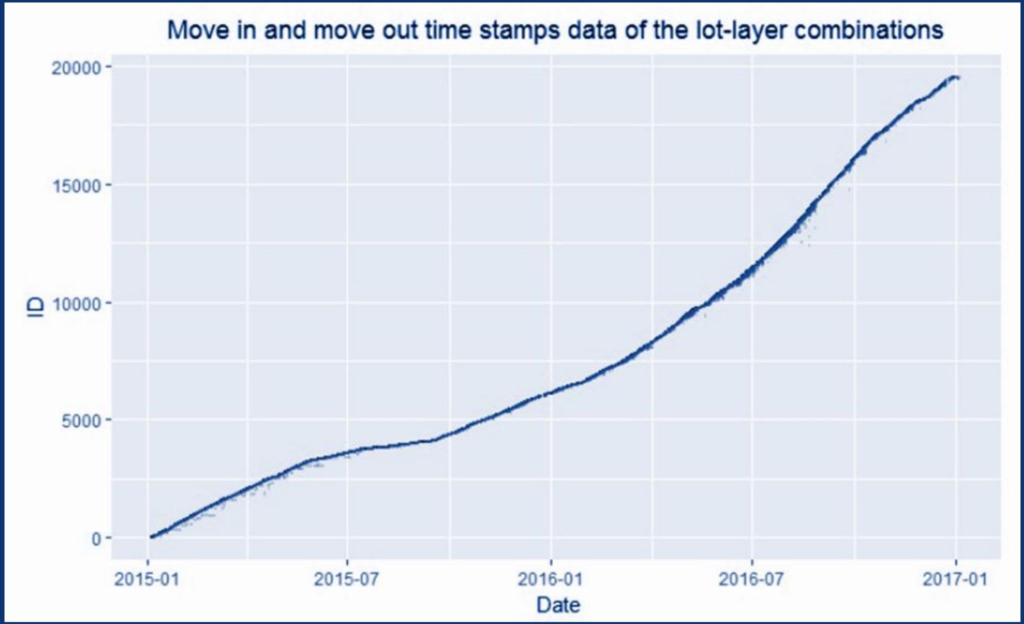


Fig 3.3-1 Timestamp data of the lot-layer combination

Customer & Machine Status are used to generate ML database which included a number of 18,532 observations about lot-layer combination

For the evaluation , the ML dataset was split up into training & testing dataset randomly with 70/30 sampling ratio. Out of the raw data , we derived 41 features that affected with lead time

The feature can be divided according to their characteristic into 2 different category
static feature : Characterized the lot
dynamic feature : reflect PD system status

3.3 DESCRIPTION OF DATA

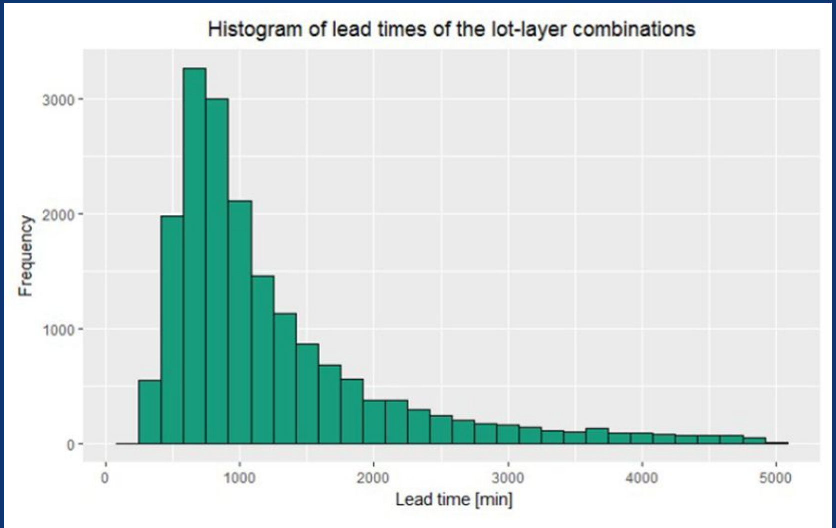
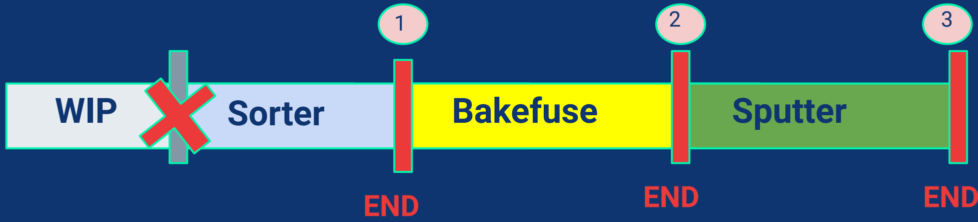


Fig 3.3-2 Histogram of the overall lead time in the analyzed steps

Lead time is calculated for the process Bakefuse & Sputter Separately . For those processes the lead time is defined as time span from the end conformation of the previous process and the observed process

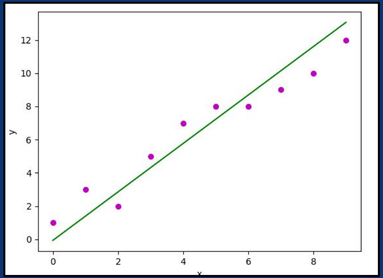
Bakefuse LT = 2 - 1

Overall LT = 3 - 1

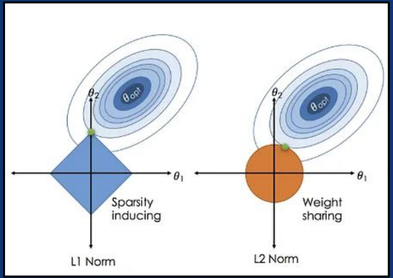


3.4 EXPLORING ML ALGORITHMS

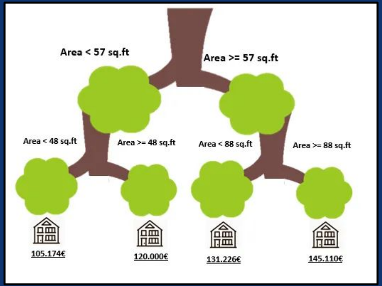
All 11 models applied in this study as shown below



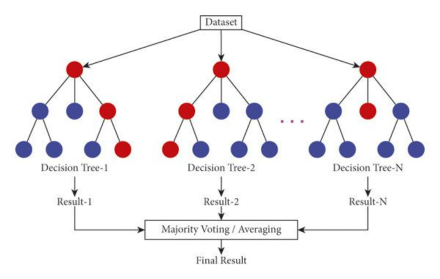
Linear Regression



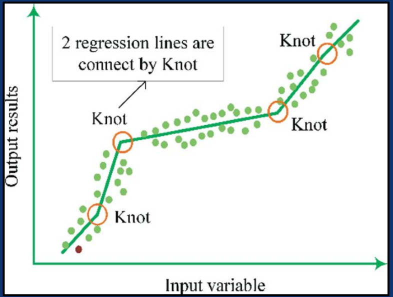
Ridge & Lasso Linear Regression



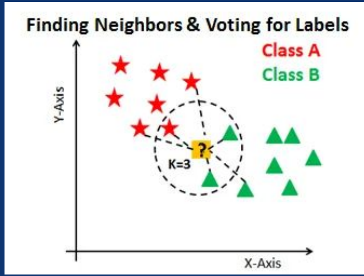
Regression Tree (Bagged RT & Boosted RT)



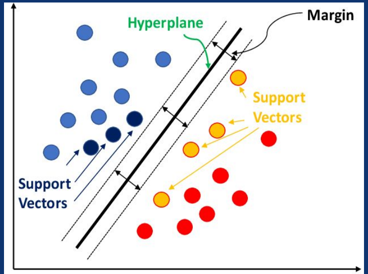
Random Forest



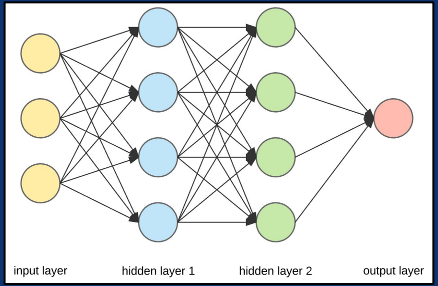
MARS



K-Nearest Neighbor



Support Vector Machine



Artificial Neural Network

3.4 EXPLORING ML ALGORITHMS - UNDERSTANDING LINEAR REGRESSION

What is Linear Regression ?

Linear Regression attempts to **model the relationship between two variables by fitting a linear equation to observed data**. One variable is considered to be an explanatory variable or an independent variable, and the other is considered to be a dependent variable

Benefits of Linear Regression

1. Ease
2. Interpretability
3. Scalability
4. Deploys and Perform well on Online Setting

ML Approaches to Linear Regression

1. Simple and Multiple Linear Regression
2. Polynomial Regression
3. Ridge & Lasso Regression
4. Decision Tree Regression (Regression Tree : RT)
5. Support Vector Regression (SVR)

3.4.1 SIMPLE & MULTIPLE LINEAR REGRESSION

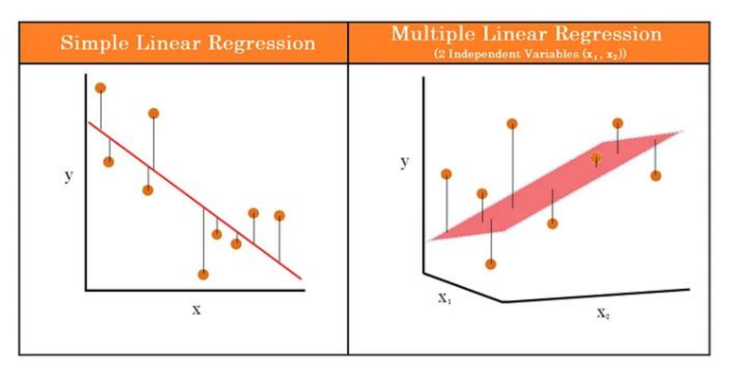
Simple Linear Regression :

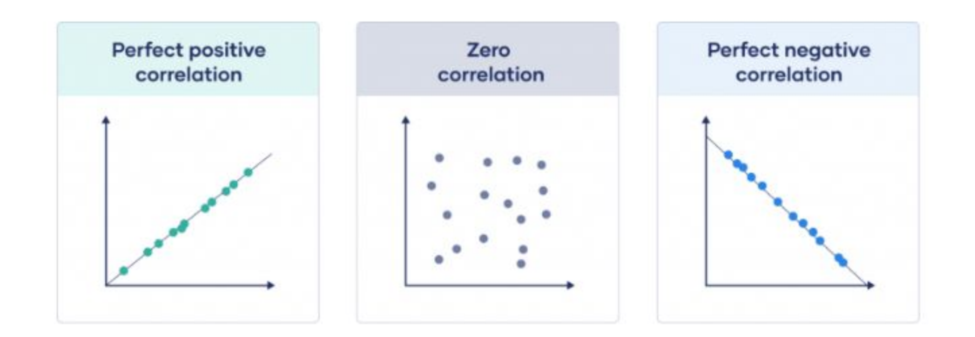
$$y = w_o + w_1 x$$

Multiple Linear Regression :

$$y = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$$

Where xi is the i-th feature with its own wi weight.





3.4.2 RIDGE & LASSO REGRESSION :

1. **Ridge Regression**

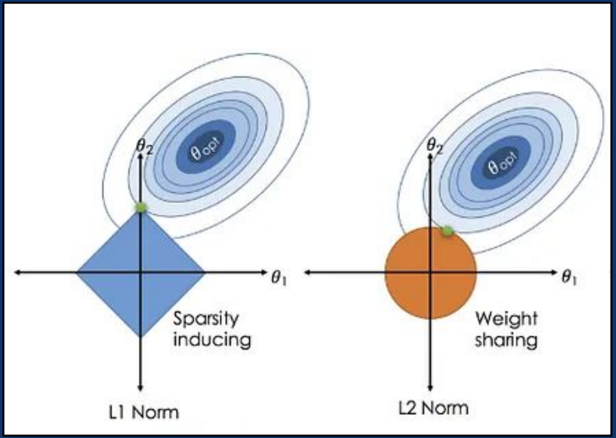
A linear regression model that implements **L2 norm** for regularisation

$$\min_{w \in \mathbb{R}^p, b \in \mathbb{R}} \sum_{i=1}^{n} (w^T \mathbf{x}_i + b - y_i)^2 + \alpha \|w\|^2$$

2. **Lasso Regression**

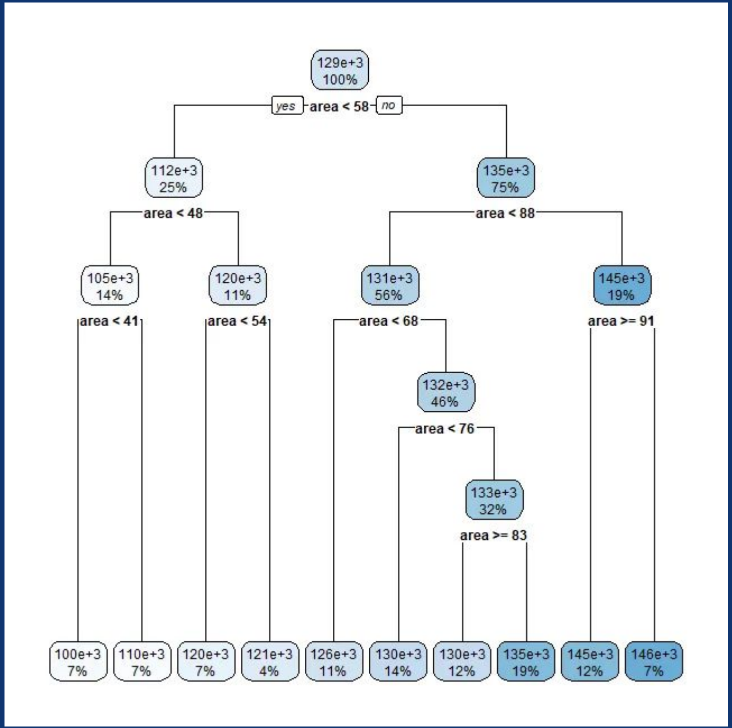
A linear regression model that implements **L1 norm** for regularisation

$$\min_{w \in \mathbb{R}^p, b \in \mathbb{R}} \sum_{i=1}^{n} (w^T \mathbf{x}_i + b - y_i)^2 + \alpha \|w\|_1$$



3.4.3 REGRESSION TREE (RT) :

A regression tree is basically a **decision tree that is used for the task of regression** which can be used to predict continuous valued outputs instead of discrete outputs.



Bagged RT

- RT + Bootstrap Aggregation
- This ensemble model is applied for Overfit Reduction

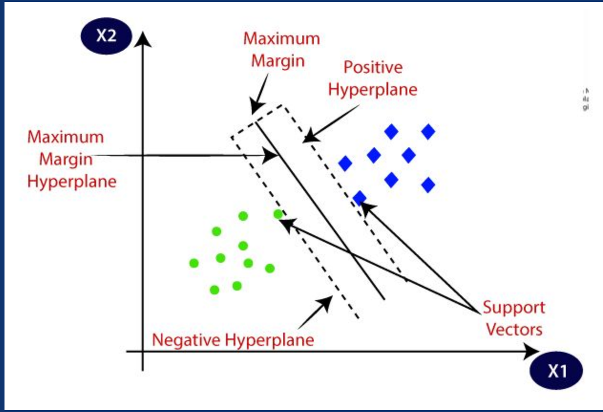
Boosted RT

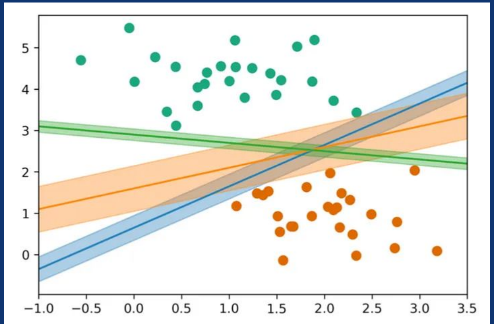
- RT + Boosting Method
- This model is applied for improving accuracy

3.4.4 SUPPORT VECTOR MACHINE (SVM) :

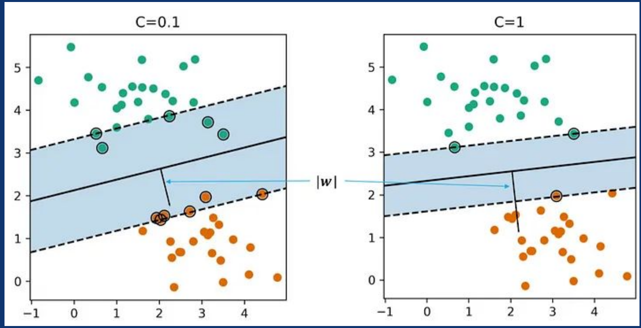
The goal of the SVM algorithm is to **create the best line** or decision boundary **that can segregate n-dimensional space into classes** so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane

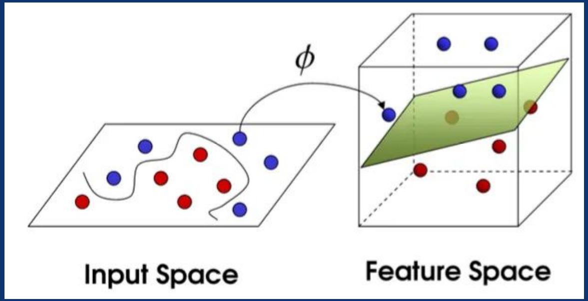




Max-Margin and Support Vector



Parameter C



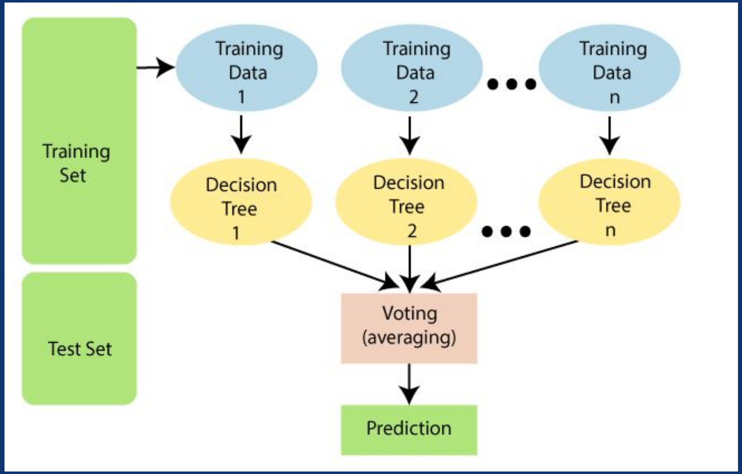
Kernels

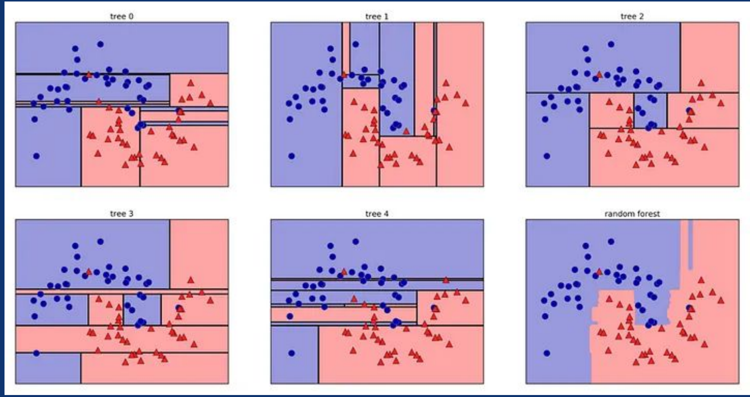
3.4.3 RANDOM FOREST (RF) :

"Random Forest is a **classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."**

Instead of relying on one decision tree, the random forest takes the prediction from **each tree and based on the majority votes** of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting



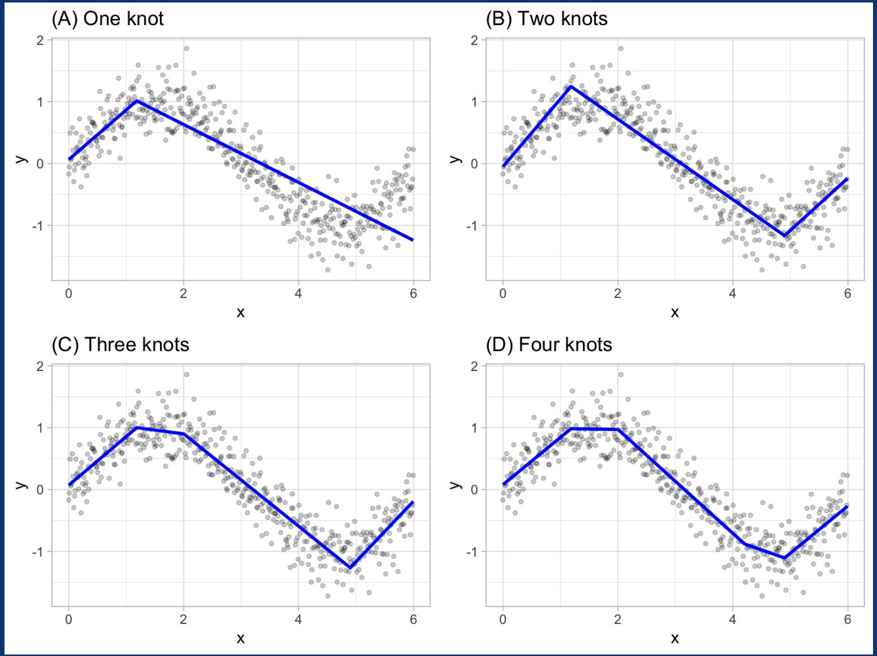


3.4.5. MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) :

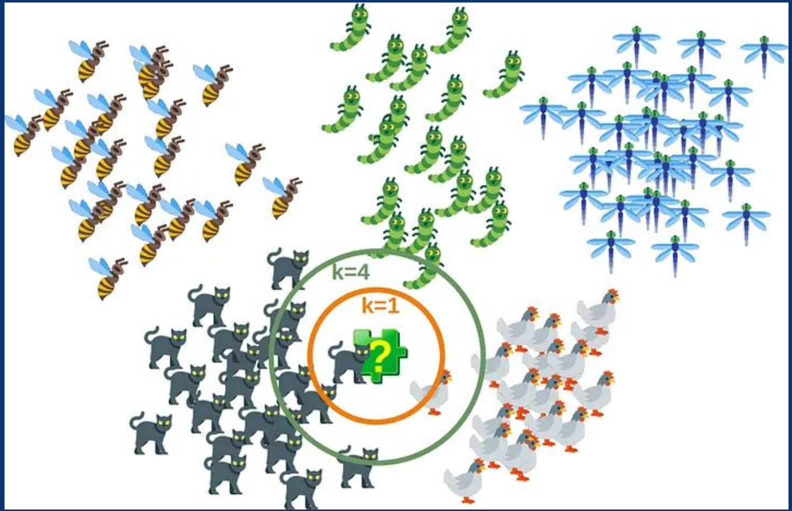
MARS - is an algorithm for complex non-linear regression problems. It **provides a convenient approach to capture the nonlinear relationships in the data by assessing cutpoints (knots)** similar to step functions. The procedure assesses each data point for each predictor as a knot and creates a linear regression model with the candidate feature(s)

Key to the MARS algorithm is how the basis functions are chosen. This involves two steps: the growing or generation phase called the forward-stage and the pruning or refining stage called the backward-stage.

- **Forward Stage**: Generate candidate basis functions for the model.
- **Backward Stage**: Delete basis functions from the model



3.4.6. K - NEAREST NEIGHBOR (KNN) :



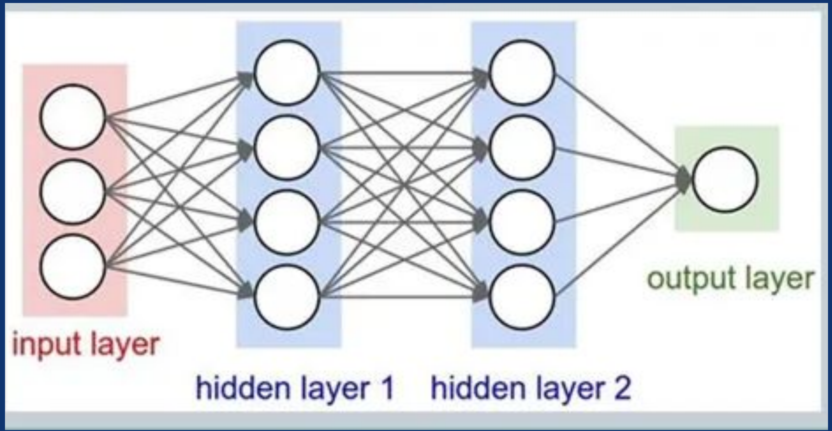
K-nearest neighbors (KNN) is a type of supervised learning algorithm used for both regression and classification. KNN tries to **predict the correct class for the test data by calculating the distance between the test data and all the training points. Then select the K number of points which is closest to the test data**.

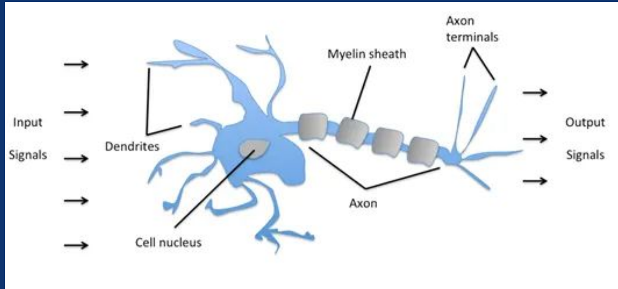
The KNN algorithm calculates the probability of the test data belonging to the classes of 'K' training data and class holds the highest probability will be selected. In the case of regression, the value is the mean of the 'K' selected training points.

3.4.7. ARTIFICIAL NEURAL NETWORK (ANN) :

Artificial neural networks are one of the main tools used in machine learning. As the “neural” part of their name suggests, they are **brain-inspired systems** that are intended to replicate the way that we humans learn

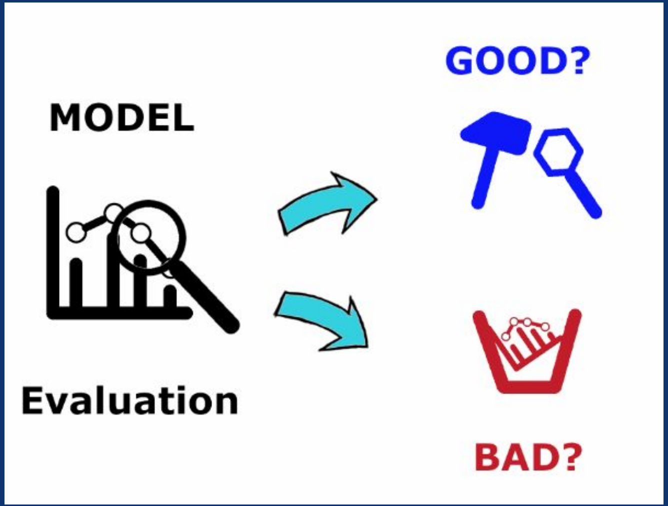
Neural networks **consist of input and output layers, as well as** (in most cases) **a hidden layer consisting of units that transform the input into something that the output layer can use.** They are excellent tools for finding patterns that are far too complex or numerous for a human programmer to extract and teach the machine to recognize.





Schematic of a biological neuron.

3.5 MODEL EVALUATION



For this study, we will be talking about 2 different ways for evaluating model below

- Error Measure
- Sensitivity Analysis

And our 11 models previously will be evaluated by using **R Programming (from R Studio)**

3.5.1 ERROR MEASURES

Model with **lowest Error** is needed !!!

During the model building and feature selection, **10-fold cross validation** was performed to estimate the prediction accuracy of the models on the independent test set. **The accuracy of the model were measured with 5 various error measures** as below

1. Mean Absolute Error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x|$$

2. Mean Absolute Percentage Error (MAPE or M)

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

3. Mean Squared Error (MSE)

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

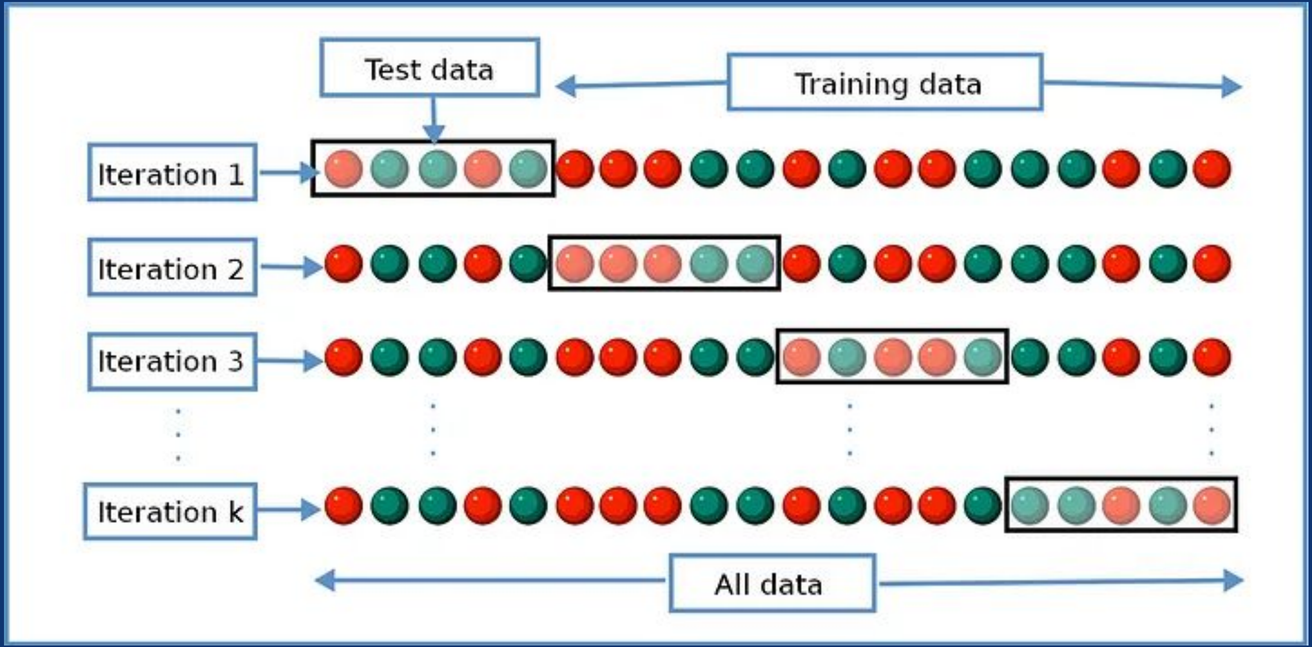
4. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

5. Normalized Root Mean Squared Error (NRMSE)

$$\mathrm{NRMSE} = 100 \frac{\sqrt[2]{\frac{1}{n}\sum_{i=1}^{N}(P_i - R_i)}}{R_{max} - R_{min}}$$

3.5.2 K-FOLD CROSS VALIDATION



3.5.3 SENSITIVITY ANALYSIS

Sensitivity analysis is a technique used to determine how changes in an input variable affect the output of system or model it is used ***to identify the most critical factor that can impact the performance*** of a system or model

General form of formula can be written as....

Sensitivity = Percentage change in output / Percentage change in input

How to interpret……

Positive (+) : Direct Relationship

Negative (-) : Inverse Relationship

Sensitivity analysis is used in the business world and in the field of economics. It is commonly used by financial analysts and economists and is also known as **what-if analysis**

TODAY'S AGENDA

PART 1
Introduction

PART 2
Literature Review

PART 3
Method Explanation

PART 4
Result Explanation

PART 5
Future Research / Recap / Q&A

4.1 COMPARISON OF ACCURACY & SENSITIVITY ANALYSIS

Table 4-1 Accuracy of the tested ML algorithms based on different error measures

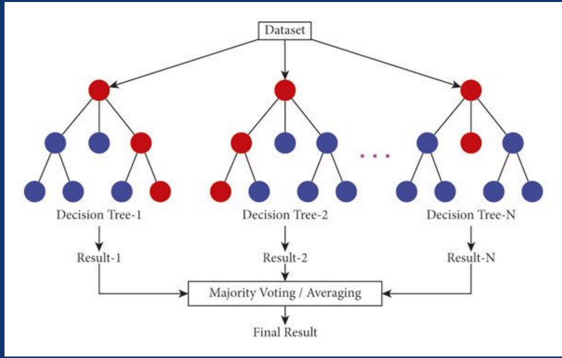
	LM	Ridge	Lasso	RT	bagged RT	RF	boosted RT	SVM	MARS	kNN	ANN
MAE	487	510	508	563	394	390	397	423	488	504	535
MAPE	42.7	45.0	44.7	53.5	33.9	33.8	33.9	30.9	43.15	44.0	53.4
MSE	529408	573520	572939	639617	369993	360780	369414	500693	513638	554897	658852
RMSE	727	757	756	799	608	600	607	707	716	745	771
NRMSE	15.2	15.8	15.8	16.7	12.7	12.5	12.7	14.77	14.9	15.5	17.3

Table 4-2 Description & Sensitivity analysis result of 10 most important variable of the models

Feature	Description	RF	boosted RT
MovDeparture	Moving average of the inter departure times of the last 20 lot-layers	2,7	3.3
ArrivalHour	Hour of the arrival time	-7,6	-13.9
WD	Weekday of the arrival time	3.6	2.7
SumMedOTs	Median of the product's lead time (in the analyzed process steps)	5.2	1.5
WIP	Work in progress: number of lot-layers in the analyzed process steps	5.3	2.6
WIPtimeBfMed	Work in progress: expected work content in minutes by process step bakefuse	2.9	1.7
SpEffPrevDay	Capacity utilization of machines in process step sputter on the previous day	0.1	1.9
MovArrival	Moving average of the inter arrival times of the last 10 lot-layers	-1.3	-1.3
medOTProdRout	Mean of median operations times of a product on a given route	3.6	2.1
SBPrevDay	Time in standby status of the machines of sputter on the previous day	1.2	0.8

4.2 CONCLUSION – WHICH MODEL?

Our Final model suggested for lead time prediction in this particular case is a *RF model with all the 8 variables with Positive Sensitivity analysis result* the original NRMSE of 12.5 is reached



Feature	RF
MovDeparture	2,7
WD	3.6
SumMedOTs	5.2
WIP	5.3
WIPtimeBfMed	2.9
SpEffPrevDay	0.1
medOTProdRout	3.6
SBPrevDay	1.2

4.1 COMPARISON OF ACCURACY & SENSITIVITY ANALYSIS

Table 4-1 Accuracy of the tested ML algorithms based on different error measures

	LM	Ridge	Lasso	RT	bagged RT	RF	boosted RT	SVM	MARS	kNN	ANN
MAE	487	510	508	563	394	390	397	423	488	504	535
MAPE	42.7	45.0	44.7	53.5	33.9	33.8	33.9	30.9	43.15	44.0	53.4
MSE	529408	573520	572939	639617	369993	360780	369414	500693	513638	554897	658852
RMSE	727	757	756	799	608	600	607	707	716	745	771
NRMSE	15.2	15.8	15.8	16.7	12.7	12.5	12.7	14.77	14.9	15.5	17.3

Table 4-2 Description & Sensitivity analysis result of 10 most important variable of the models

Feature	Description	RF	boosted RT
MovDeparture	Moving average of the inter departure times of the last 20 lot-layers	2,7	3.3
~~ArrivalHour~~	~~Hour of the arrival time~~	~~7,6~~	~~13.9~~
WD	Weekday of the arrival time	3.6	2.7
SumMedOTs	Median of the product‘s lead time (in the analyzed process steps)	5.2	1.5
WIP	Work in progress: number of lot-layers in the analyzed process steps	5.3	2.6
WIPtimeBfMed	Work in progress: expected work content in minutes by process step bakefuse	2.9	1.7
SpEffPrevDay	Capacity utilization of machines in process step sputter on the previous day	0.1	1.9
~~MovArrival~~	~~Moving average of the inter arrival times of the last 10 lot-layers~~	~~-1.3~~	~~-1.3~~
medOTProdRout	Mean of median operations times of a product on a given route	3.6	2.1
SBPrevDay	Time in standby status of the machines of sputter on the previous day	1.2	0.8

TODAY'S AGENDA

PART 1
Introduction

PART 2
Literature Review

PART 3
Method Explanation

PART 4
Result Explanation

PART 5
Future Research / Recap / Q&A

5.1 FUTURE RESEARCH AGENDA

- **Using this approach with those 8 variable is OK → Extend to other process and Whole Production System**

- **Variable & Learning Algorithm are suitable → Extend to other Industries**

- ***Feature Codebook* is needed → to be Guideline for tuning existed feature & defining and testing new feature for this process & other different process type**

- **Need more study in Analyzing *interrelation between variable with negative sensitivity*** as well

5.2 RECAP

- Traditional planning & Control method mostly calculated **Average Lead Time** (which derived from MES)
- That point results in the **deficiency of PPS** (including Quality & Efficiency) – ***More Error → Bad Performance***
- **This study** try to propose new idea for improving lead time Prediction method in complex system like semiconductor manufacturing by **using Supervised ML method**
- 3 process steps in this study : **Sort / Bakefuse / Sputter**
- Our used ML database are generated by using Machine & Customer data which includes number of **over 18K observations** about **lot-layer combination** with 41 feature (35 Num & 6 Cat)
- **11 Different Models** are evaluated by **5 different Error measures & Sensitivity Analysis** → **Our best model is RF with last 8 features which have NRMSE = 12.5**
- This improved model will be studied with other similar process & whole production system or even may be used for other industries as well

5.3 FROM MY VIEW

5.4 REFERENCE

- https://www.sciencedirect.com/science/article/pii/S2212827118303056
- https://www.deskera.com/blog/manufacturing-lead-time/#significance-of-manufacturing-lead-time
- https://byjus.com/jee/semiconductors/
- https://www.modify.in.th/20799
- https://www.slri.or.th/bl6a/process/film-deposition/12-process/44-sputtering.html
- https://kamboonchob.medium.com/%E0%B9%80%E0%B8%82%E0%B9%89%E0%B8%B2%E0%B9%83%E0%B8%88-crisp-dm-%E0%B8%89%E0%B8%9A%E0%B8%B1%E0%B8%9A%E0%B9%80%E0%B8%A3%E0%B9%88%E0%B8%87%E0%B8%A3%E0%B8%B1%E0%B8%94-b0913050198f
- https://medium.com/@pradyasin/random-forest-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-74d2a0af3d7
- https://medium.com/@pradyasin/support-vector-machines-svm-943f9a732a69
- https://1stcraft.com/the-component-of-production-system/
- https://pongtorn-elec.blogspot.com/2017/10/wafer.html
- https://www.sumipol.com/knowledge/what-is-production-lead-time/
- https://www.slri.or.th/bl6a/process/film-deposition/12-process/44-sputtering.html
- https://www.investopedia.com/terms/s/sensitivityanalysis.asp

THANK YOU

FOR YOUR ATTENTION