Manpower Forecasting And Allocation In The Manufacturing Sector

IE3100R System Design Project (AY2022/2023)

Group 5: Liang Tim Tien| He Zhenyu| Yab Yi Tong| Nguyen Phuong Minh

Academic supervisor: A/Prof Ye Zhiseng

Industrial supervisor: Jeslin Goh, Daniel Leow

Overview

Company Background

Infineon Technologies is among the 10 largest semiconductor manufacturers globally. Infineon Singapore, develops the local electronics market and talent pools. It is also a key point for global distribution and the major microelectronics R&D center in Asia, as well as the reference site for advanced backend manufacturing.

Smart manufacturing enabled Infineon to cut direct labour costs by 30%, improve capital efficiency by 15%, and enhance quality from defects from per million to per billion.

Problem Definition

This section investigates Infineon approach to estimate resources and time needed for NPI projects and its methods to monitor and plan project schedules. At Infineon, each project is defined by its type and class. This project aims to conduct data analysis on theoretical relationship between project specification and man hour required for each project, and formula forecast models to systemize the man hour allocation process for their projects.

Man Hour Prediction for NPI Projects

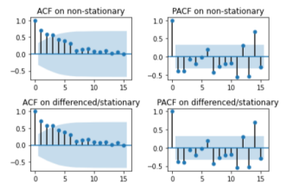
After studying effects of project specifications and RCM's biometric data on man hour needed in a project, the team is to predict monthly working hours required for NPI projects for the next 12 months

Integration of NPI Project Resource to Tableau Dashboard

Currently, project information is scattered across various database that lack robustness in sharing information with different stakeholders. Infineon aims to integrate all project data onto a Tableau dashboard to improve its quality of life and project management capabilities.

Key Findings from Data Exploration

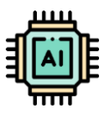
- ❑ Data Integrity
  -- > Significant amount of Missing Data
- ❑ Time Series Decomposition of Past Monthly Man Hours



- ❑ Other Project Features and Parameters
  - ➢ Distribution of Project Status
  - ➢ Distribution of Project Classes
  - ➢ RU Managers Experience and Assignment
  - ➢ Possible Missing Data for Some Managers
  - ➢ Project Types Distribution

Skill Set

Project Management
- Deliverables and Timeline
- Stakeholder engagement



Machine Learning
- Used Python programming language to build the model
- Time Series Decomposition
- Machine Learning Model

Dashboard Management
- Used Tableau to create interactive dashboard
- Use historical data

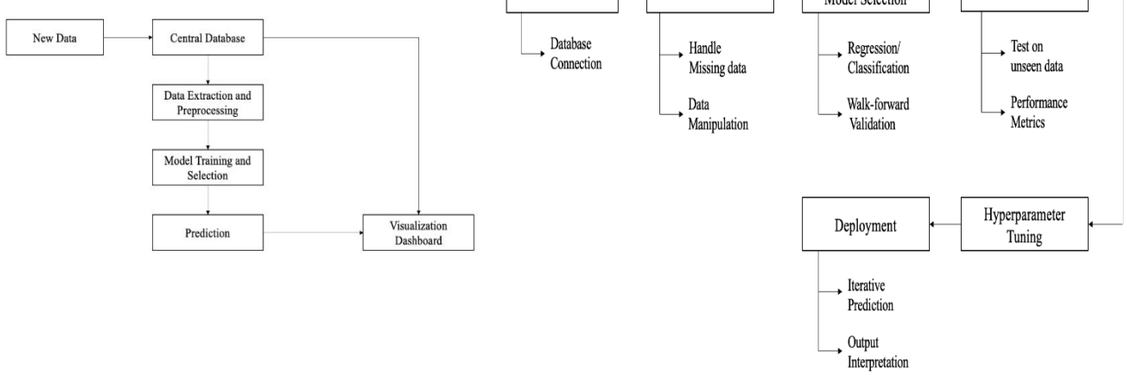
Creative Problem Solving
- Applied problem solving to tackle real world problems

Methodology

Proposed solution aims to produce a dynamic modelling function that can process data, adapt to any number of variables and output the best model. At the beginning of each month, new data from previous month's production will be extracted and fed into the predictive model to forecast future man hours required for each project and RMC managers.

In Summary, the proposed solution consists of 2 main tasks:

- ❑ Develop predictive models to forecast the required man hours for each project and the man hours each CRM manager will need to handle their monthly workloads.
- ❑ Design and create a centralised dashboard with visualization tools to facilitate the dissemination and planning of projects



Model Building & Result

Prediction Model

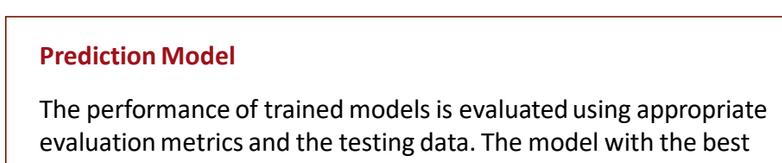
The performance of trained models is evaluated using appropriate evaluation metrics and the testing data. The model with the best performance is selected for deployment.

Data Pre-processing

Consists of 3 steps – fill in missing data, feature engineering, encoding categorical variables. To handle missing data, method used are: Fill in with median of data set, Fill in with average values of all project with similar project type and class, Fill in with last available data from the project or last available data from project with same project type, class and last milestone

Machine Learning Model

The deployment of the model involves a three-step process that include Model Selection, Hyperparameter Tuning, and Model Deployment. Since the dataset used in our project has limited scope and size, the main focus for selecting an appropriate model will be based on performance metrics, such as those obtained through validation techniques. The performance metrics of the model selection is based on Mean Absolute Percentage Error (MAPE). MAPE is chosen because MAPE places more weight on under-estimates than on overestimate.



Time Series Model

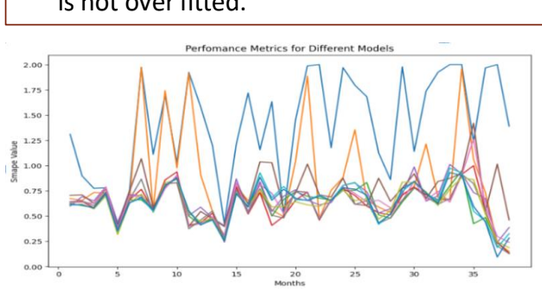
The team has considered using time series forecast model in addition to ML model. ARIMA and SARIMA models are trained with historical man hours data to estimate parameters and make predictions for future values of time series.

ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) is a time series forecasting model widely used in statistics and econometrics. It is a class of linear model that considers autocorrelation and stationarity of time series data. The ARIMA model consists of 3 components: p, d, q

ARIMA and Seasonality

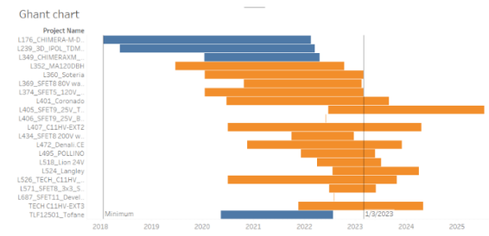
Seasonality in ARIMA models refer to presence of repeating patterns or cycles of fixed lengths within time series data. In order to model the seasonality of a time series, seasonal differencing is applied to the data. Akaike Information Criterion (AIC) is used as part of performance metrics to measure the goodness of fit of the models. The models are also validated with in-sample CV and out-of-sample data to ensure the model is not over fitted.



Models	CV SMAPE(%)
Linear Regression	0.849
Lasso	0.615
Ridge	0.613
KNN Regressor	0.671
Decision Tree Regressor	0.777
Random Forest Regressor	0.646
Gradient Boost Regressor	0.610
SVR (Linear)	0.618
SVR (Kernel)	0.635
SVR	0.627
ARIMA Model	17.097

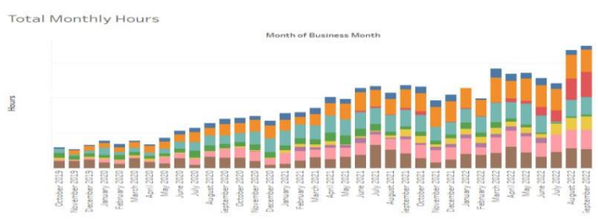
Tableau Integration

- ➢ **Aim to provide insights into project progress, project duration, manager workload, project classification over time**
- ➢ **New dashboard would integrate NPI project information from various databases as a powerful visualization tool to streamline data analytics process for various stakeholders**
- ➢ **Gantt chart was used to track progress of current project**



- ➢ **Line graph was used to represent total hours each manager works on projects over a certain period**

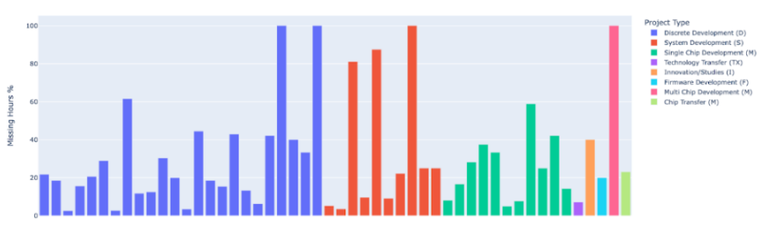
- ➢ **Stacked bar graph was used to represent total combined hours of all projects over time.**



Limitation

Limitation of Prediction Model

1. Missing Data for Target Variable
2. Lack of Confidence Interval for Prediction result



Limitation of Tableau Dashboard

1. It can only observe trends between project managers, project schedule, project class or project type
2. It will not be helpful in identifying the causes or reasons that cause overdue or overallocation or under allocation of work

Achievement

- ➢ Results from both Data Exploration and Tableau Integration are exceeding expectation and have received positive feedback from RCM Managers in Infineon. The aforementioned results are expected to be integrated into their project monitoring systems
- ➢ While the predictive model has not been deployed in Infineon's system, and trial tested, the good in-sample walk forward validation performance has shown significance as a prerequisite of a feasible model in NPI man hour forecast
- ➢ This project is a proof of concept and a stepping stone for the internal team at Infineon to develop their project management capability.

Recommendation

- **Restructure of Data Collection Strategy**
- **Bootstrapping Technique in Prediction Models**
- **Monte Carlo Simulation**
- **Establish Prediction Interval Through Ensemble Forecasting**
- **Real Time Update Feature in Tableau Dashboard**
- **Implementation of Query Feature in Tableau Dashboard**

+ableau Excel DATA COLLECTION