Predictive Analysis of Chip Retest for Semiconductor Final Test

IE3100R Systems Design Project | Group 19 | Department of Industrial Systems Engineering and Management

National University of Singapore

Group Members:
Chan Tze Kin, Chen Xiaolin
Lim Li Yi Stephanie, Xu Mingwei

Infineon Supervisors
Mr Ng Bo Yan
Mr Hee Boon Kiet

SDP Supervisor:
Associate Professor Ye Zhisheng

Project Objectives

- Prove that retest redundancy is predictable
- Analysis of historical test data and build predictive model
- Reduce retest redundancy & save company resrouces

Key Skillsets

- **Programming skills** Python, SQL and Excel
- **Data Analysis** Collection and cleaning of data
- **Modelling** Statistical analysis and machine learning modeling

Semiconductor Manufacturing

In semiconductor chip manufacturing, chips go through a vigorous testing process before they can be shipped out to customers. This is known as the final test stage in the manufacturing process, and it ensures that the customers receive chips that are of high reliability.

However, this testing process has potential for error which results in the false rejection of chips. False rejection of chips could happen for a variety of reasons not limited to issues related to test equipment. Hence, retest procedures were introduced to help investigate suspected false rejections and increase yield.

Reduction of False Rejection → Increase of yield

Problem Description

False rejection of chips lowers the manufacturing yield for the company resulting in loss due to unsold and disposed products. At the same time, it is unreasonable to retest every lot of failed chips as retesting lots can be costly in terms of time and resources for the company. Hence, it is important to be able to selectively send lots to retest in order to maximise recovery and minimise unnecessary testing.

Excessive Testing | Outright disposal

Methodology

Data Labelling

Historical test data was labelled redundant based on 4 criteria:
1. Same stop reason
2. Same lot number
3. Same test mode
4. Different stop time

Label Validation

Random samples of cases labeled as redundant were analysed to validate the accuracy of the labelling

Recovery Analysis

Recovery rates from one test to the next retest were calculated and statistically analysed

Data Cleaning

Correctly labeled data was cleaned using 4 main tools:
1. Missing data filtering
2. Low variance filtering
3. Pearson's correlation
4. Chi-square hypo. testing

Modelling

Four machine learning models were attempted:
1. Logistic Regression
2. Support Vector Machine
3. Random Forest
4. Gradient Boosting

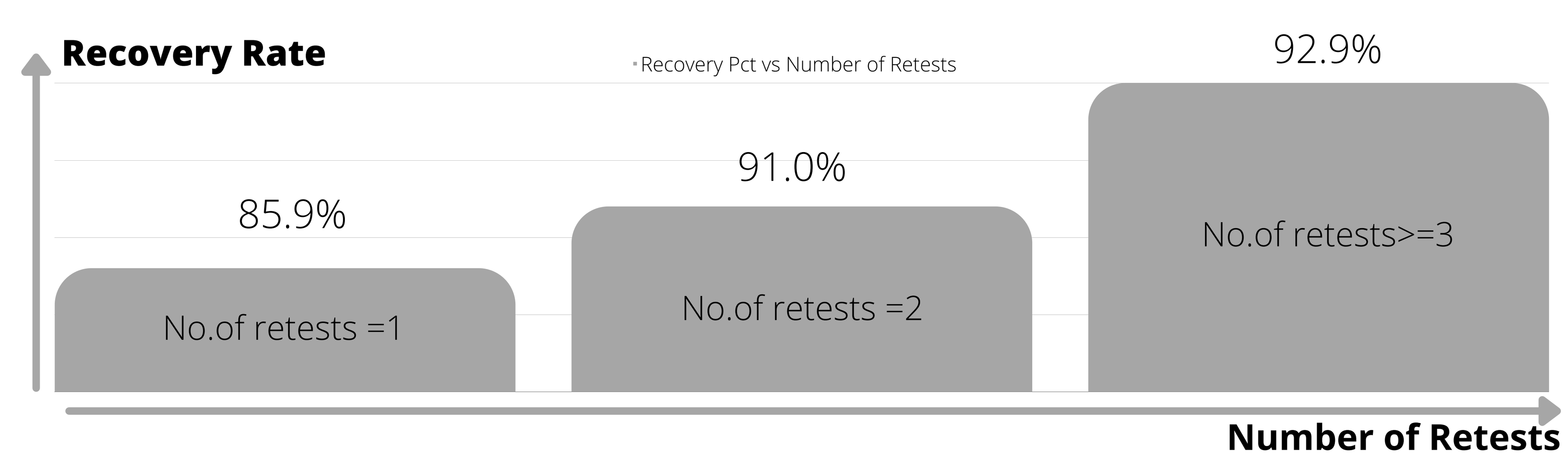
Label Validation

After further investigation, the initial 4 criteria were deemed insufficient to label redundancy. Several issues were observed and subsequently, the labelling criteria was updated

- Initial 4 criteria
- Recovery rate below 40%
- Data-issue free
- Handling-issue free

Recovery Analysis

The group has found that generally, when there is higher number of retests, the recovery rate increases. This provides a general guideline, proving that more retests could improve yield to certain extent though this Is only at an aggregate level.



Data Cleaning

Missing data filtering
- Variables with > 80% missing data were removed
- Variable means were filled for the rest of numerical variables

Pearson's Correlation
- Highly correlated numerical variables were identified
- Variable within each pair of the highly correlated variables was randomly selected and removed.

Low Variance Filtering
- All numerical variables were first normalized
- Numerical variables with a variance of lower than 0.03 were removed

Chi-square hypo. testing
- P-values were computed between each categorical variable and the target 'Redundancy'
- Variables with significantly low correlation with the target were removed

→ **Cleaned Data**

Results

Performance Metrics

Selection Criterion: Precision = TP / (TP + FP) as the most important criterion

Reasoning: High precision indicates a low false-positive rate. Under this project, false-positive indicates that the group predicted the chip-retest to be redundant where actually the retest is not redundant. From a cost perspective, wrongly discarding a lot of chips is more costly than conducting additional rounds of retesting.

Final Model Selection: Random Forest (Overall the best)

Model Performance

Logistic Regression
Precision: 0.5197
Recall: 0.6667
F1-score: 0.5841

Support Vector Machine
Precision: 0.5556
Recall: 0.4798
F1-score: 0.5149

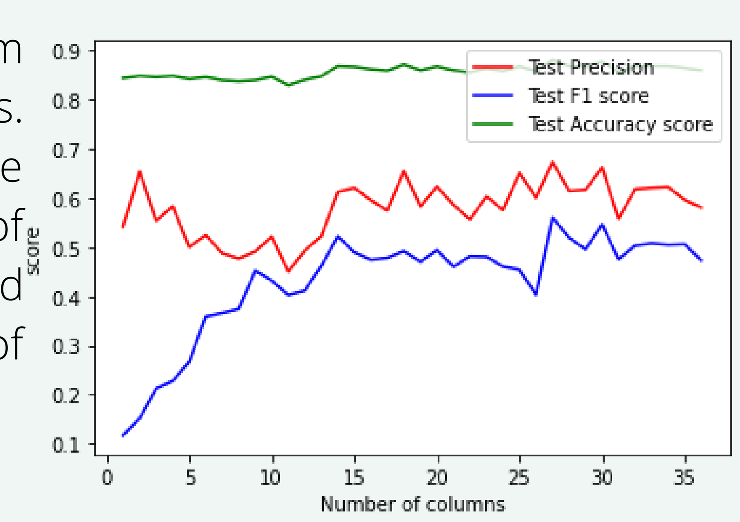
Random Forest
Precision: 0.6738
Recall: 0.4798
F1-score: 0.5605

Gradient Boosting
Precision: 0.6118
Recall: 0.5253
F1-score: 0.5652

Feature Selection

Since there are over 50 columns, overfitting is likely to occur if the model includes every variable. As such, Select-K-Best algorithm was adopted for four machine learning models. By generating model scores with respect to the number of variables, an optimal number of variables with high scores could be determined without including an excessive number of variables.

Ex. Random Forest Feature Selection Performance



Recommendations and Future Study Directions

- The current predictive model serves as a useful guideline to predict redundancy
- To improve the existing model, a dataset of a larger scale should be chosen to avoid issues such as imbalanced data and limited sample size for selected basic types
- A more robust machine learning model could be chosen to reap the global maximum score during Hyperparameter Tuning stage