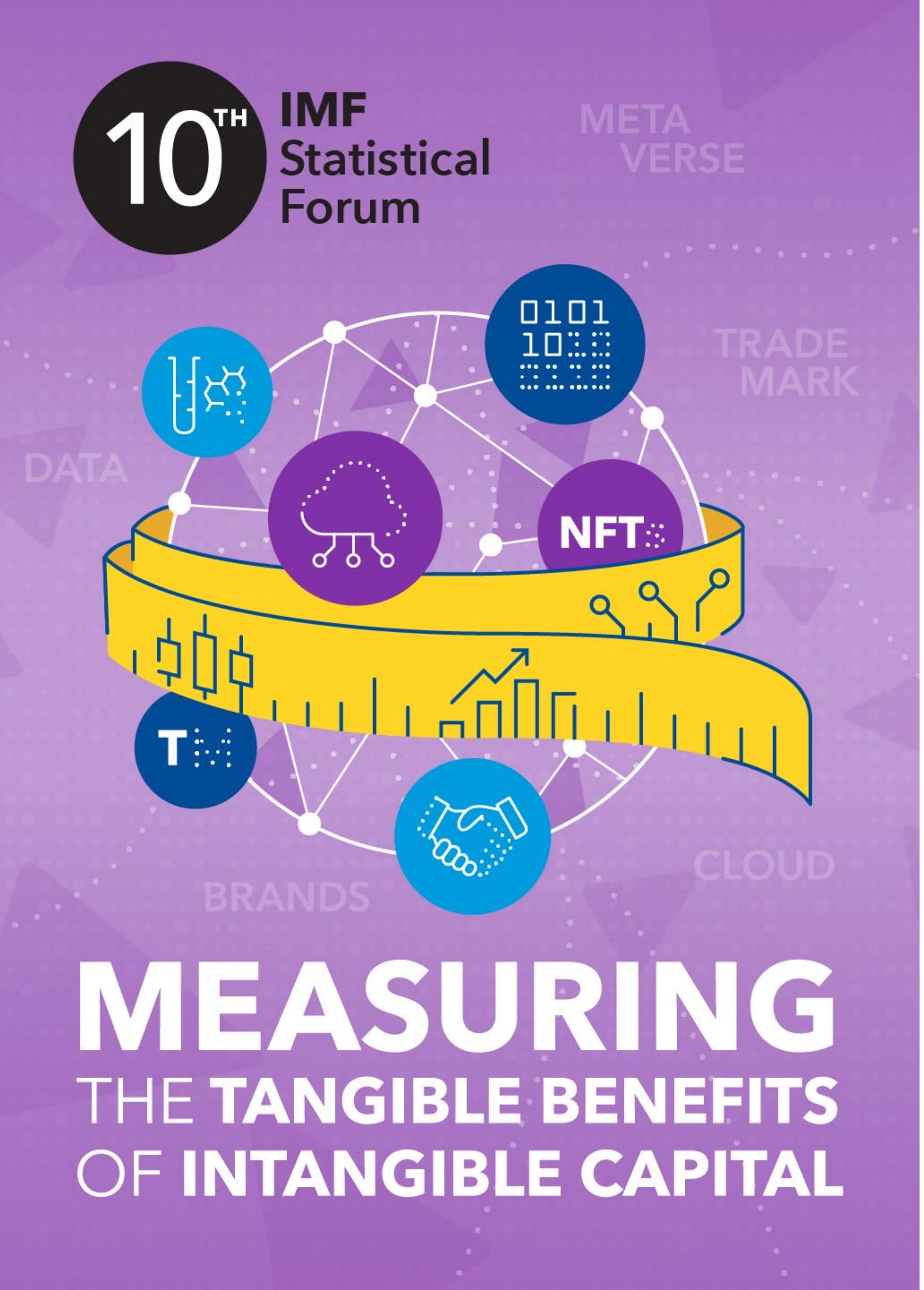


Measuring Intangible Assets Using Parametric and Machine Learning Approaches

NOVEMBER 16-17, 2022

Atika Nashirah Hasyyati, Adhi Kurniawan
Indonesia

Definition of Intangible Capitals

- Intangible capital is defined as assets that have no physical or financial embodiment (Alsamawi et al., 2020).
- Corrado et al. (2005) classified intangible assets into three categories:
  - computerised information
  - scientific and creative property
  - economic competencies.
- Vosselman (1998) classified the intangible investment into two components: core components and supplementary components.

The core components:

- research and experimental development (R&D)
- education and training
- Software
- Marketing
- mineral exploration
- licenses, brand, and copyright
- patents.

The supplementary component:

- development of organization
- engineering and design
- construction and the use of databases
- remuneration and innovative ideas
- other human resource development (training excluded).

Research objectives

- Intangible capital as the results of digitalization and globalization has not been fully measured yet in the economy because of several challenges.
- The limitation of data sources and the methodological issue related to how to measure and to capitalize intangible asset are some fundamental issues
- Research objectives:
  - Studying the association between intangible capitals and the business performance using survey data
  - Examining the accuracy of intangible capital variables to predict business performance using some statistical methods.
  - Utilizing google review data to investigate the potential effect of information in the google review for branding.

Data Sources

Bussiness Characteristics Survey

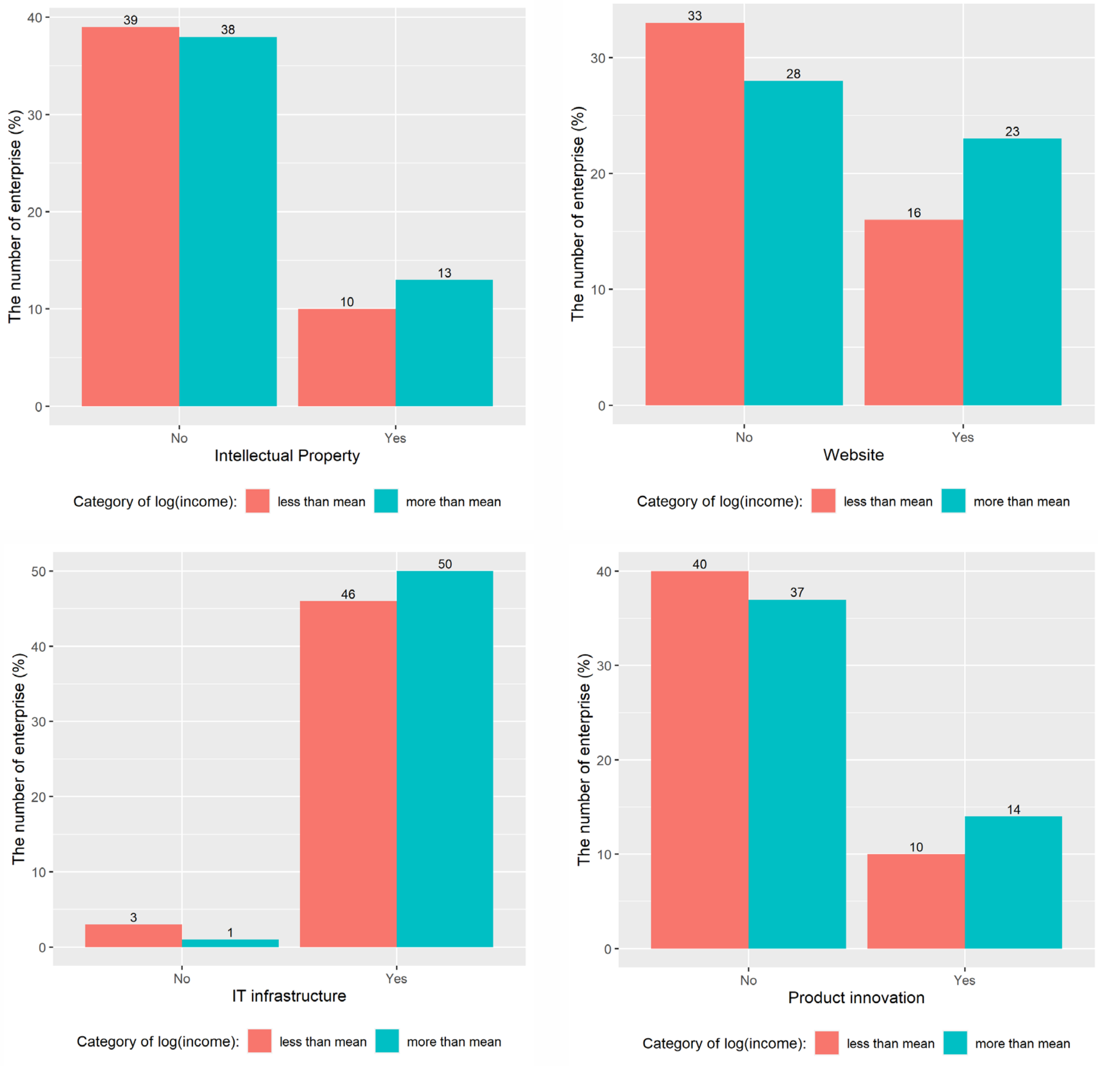
- Annual survey
- Probability sampling
- National level estimate
- Target sample size = 8,300 medium and large enterprise
- Face to face interview

Google reviews data

- Focusing on google review of accommodation / hotel
- Gathering data using Google Places API ($200 free for the first month of using Google Cloud Platform)
- Data wrangling
- Deterministic matching

Measuring Intangible Capital based on Survey Data

Descriptive Statistics



- The medium and large enterprise using intangible capitals tend to have higher income compared to the enterprise that do not utilize intangible assets.
- For enterprise using intangible capital, the number of enterprise having high income is larger than the number of enterprise having low income.
- On the contrary, in case of enterprise that do not use intangible capital, the number of enterprise having low income is larger than the number of enterprise having high income.

Measuring Intangible Capital based on Survey Data

Parametric Approach

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon$$

y : income

x_1 : the number of employees

x_2 : foreign investment (0=No, 1=Yes)

x_3 : intellectual property (0=No, 1=Yes)

x_4 : IT infrastructure (0=No, 1=Yes)

x_5 : the use of website for branding (0=No, 1=Yes)

x_6 : product innovation (0=No, 1=Yes)

IT infrastructure, the use of website for branding and product innovation are statistically significant in the model, while intellectual property is not significant

Survey weight is taken into account for building model since the data is collected by using *unequal probability sampling*.

Variable	Estimate	Standard Error	t.value	p-value
Intercept	20.201	0.139	145.63	0.00
The number of employees	0.001	0.000	1.59	0.11
Foreign investment	1.878	0.198	9.48	0.00
Intellectual property	0.055	0.075	0.73	0.47
IT infrastructure	**1.197**	**0.144**	**8.32**	**0.00**
Website for branding	**0.597**	**0.066**	**9.02**	**0.00**
Product innovation	**0.360**	**0.077**	**4.69**	**0.00**

Measuring Intangible Capital based on Survey Data

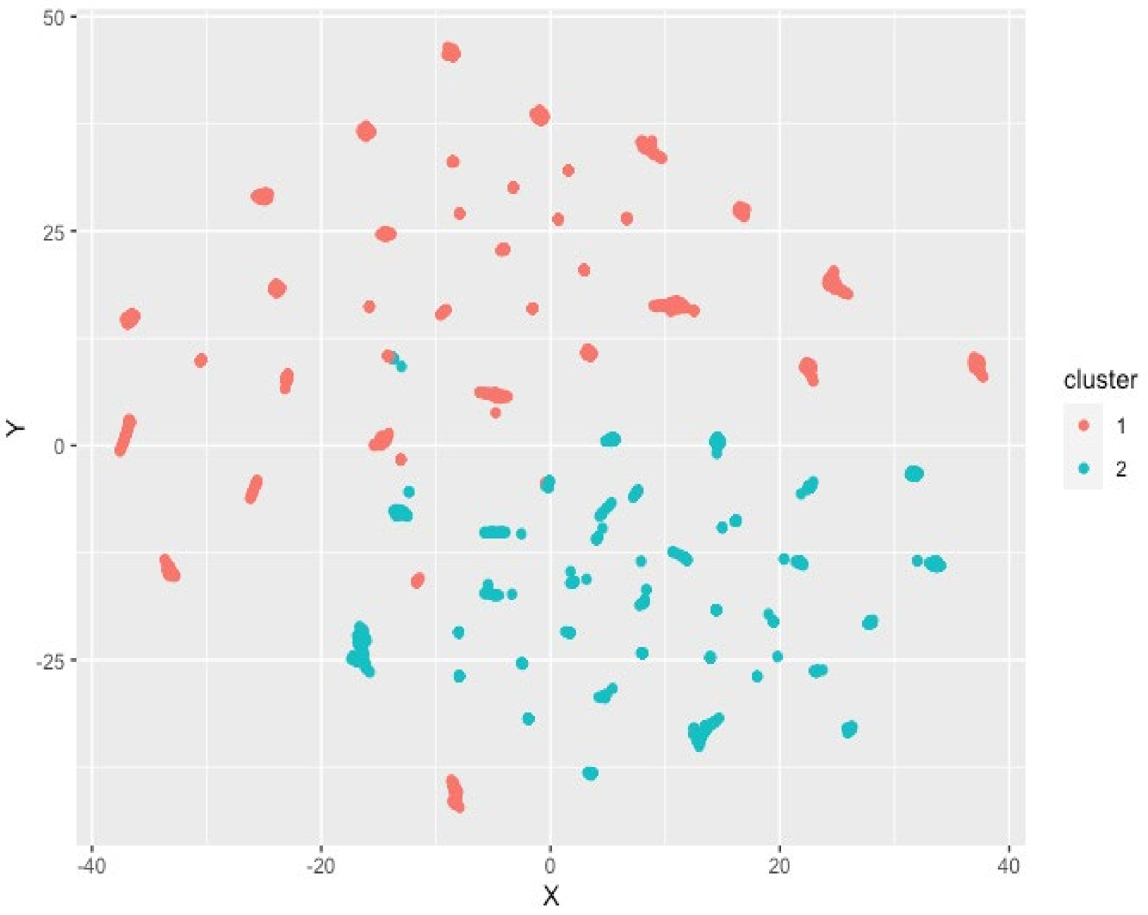
Prediction of log(income) based on six independent variables

No	Method	RMSE	R-Square	MAE
1	Parametric regression	2.3508	0.0854	1.7528
2	KNN-regression	1.9640	0.3224	1.4646
3	Bagging	1.9502	0.3293	1.4461
4	Random Forest	1.9500	0.3418	1.4519
5	**XGBoost**	**1.9331**	**0.3471**	**1.4347**

- Prediction based on machine learning approach tend to have higher accuracy than parametric approach.
- Predicting log(income) using XGBoost results in the lowest Root Mean Square Error (RMSE), the highest R-square, and the lowest Mean Absolute Error (MAE).
- There is just a slight difference between the accuracy of KNN-regression, Bagging and Random Forest.

Measuring Intangible Capital based on Google Reviews Data

- An unsupervised machine learning method is used to get categories of branding based on the Google Reviews data.
- Six variables used: (1) the availability of photo, (2) rating, (3) the availability of information about facilities, (4) the number of reviewers, (5) opening hour information, (6) province.
- Implementing the concept of Gower distance with Partitioning Around Medoids (PAM)
- Using Silhoutte width for selecting the number of cluster → the number of clusters is two clusters according to the highest Silhouette width.
- Applying t-distributed stochastic neighbourhood embedding (t-SNE) to visualize the cluster in two dimensional space



Most of observations that are in the first cluster provide information about facilities, photos, and opening hours with the average of rating is 4.3. Most of observations in the second cluster does not provide information about opening hour with the average of rating is 4.0

Measuring Intangible Capital based on Google Reviews Data

Prediction

- In order to examine the reassure that those variables are good predictors for branding category, some supervised learning methods are performed, such as regression trees, random forest, neural network, XGBoost, and penalized multinomial regression.
- All methods result in same conclusion that the accuracy are very high (around 99%).

No	Method	Accuracy	Kappa
1	Regression tree	0.9938	0.9875
2	Random forest	0.9977	0.9953
3	Neural network	0.9969	0.9937
4	**XGBoost**	**0.9992**	**0.9984**
5	Penalized multinomial regression	0.9984	0.9968

Measuring Intangible Capital based on Google Reviews Data

Association between rating and the number of reviewers

We can estimate the average of rating at province level using:

$$I_i = \frac{P_{ij} I_{ij}}{\sum_j P_{ij} I_{ij}}$$

Based on the google reviews data and official data, the ratio of reviewers to visitors is estimated by:

$$r_i = \frac{\frac{N_i}{n_i} \sum_j P_{ij}}{O_i T_i}$$

I_i : the rating average at *i*-th province

P_{ij} : the number of reviewers at *i*-th province *j*-th hotel

r_i : ratio of the number of reviewers to the number of visitors at *i*-th province

O_i : occupancy rate at *i*-th province (based on official survey data)

T_i : the total number of hotel room at *i*-th province (based on official survey data)

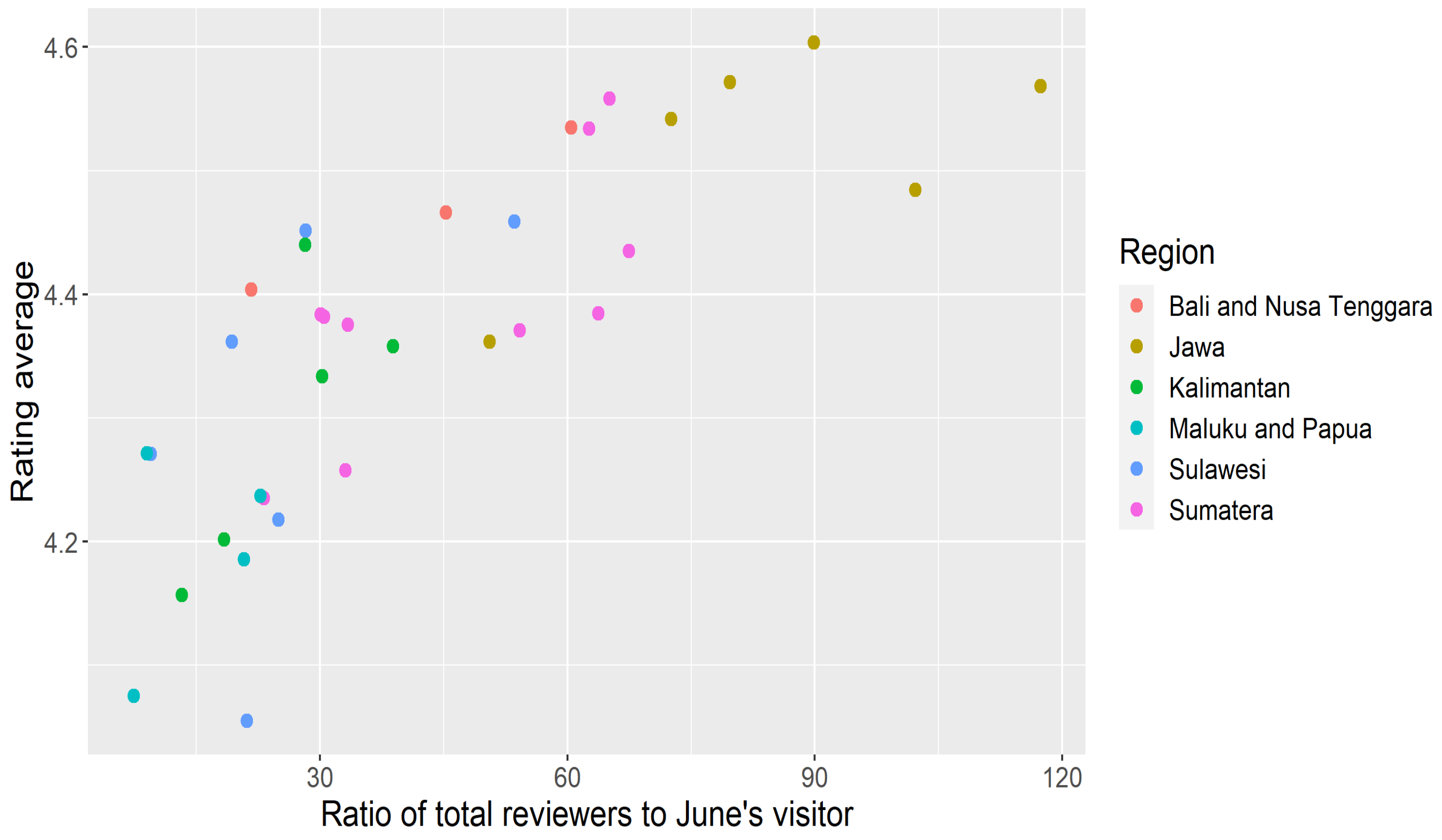
N_i : the total population of hotel at *i*-th province (based on official survey data)

n_i : the number of hotel in the dataset (based on google review data) at *i*-th province

Measuring Intangible Capital based on Google Reviews Data

Association between rating and the number of reviewers

- There is positive linear correlation between rating and the number of reviewers with Pearson correlation coefficient approximately 0.78.
- The higher the number of reviewers, the higher rating average of the province.
- It also indicates that visitors who wrote the review tend to be the visitors who had good experience when staying in the hotel.
- The tendency of hotel visitors to give reviews in Jawa (Java) is higher than those in other island.



Conclusion and Recommendation

Conclusion:

- The results of parametric regression show that the proxy of intangible capitals used in this paper, except intellectual property, has significant contribution on the business performance.
- Based on machine learning approach, variables that are obtained from google reviews can be used to predict the use of branding with the high accuracy.

Recommendation:

- The combination of using official survey data and other source of data (e.g. Big data) in data analysis through statistical data integration should be continuously developed to measure intangible capital.