Integrating data mining technique and AHP in market analysis to propose new product development in real estate

Yunita¹, M. Galinium¹, Lukas²

¹Faculty of Engineering and IT Swiss German University, EduTown BSD City, Tangerang, INDONESIA
²Faculty of Engineering, Universitas Katolik Indonesia Atma Jaya, Jakarta, INDONESIA

E-mails: unita.st@gmail.com, maulahikmah.galinium@sgu.ac.id, lukas@atmajaya.ac.id

Abstract. New product development in real estate industry is a challenging process since it is related to long term concept and high cost. A newly proposed product development should meet customer need and their preferences which appropriate with customer buying power and company value. This research use data mining for profiling customer transaction and Analytic Hierarchy Process (AHP) method for product selection in new product development. This research utilizes Weka as data mining open source software to profiling data customers. The analysis correlated product preferences and profiling demography such as city, age, gender and occupation. Demography profiles gives description buying power and product preferences. The products proposed are based on customer profiles and rank of the product by AHP method. The product with the highest score will be proposed as new product development. Case studies of this research are real estate projects in Serang, Makassar, and Balikpapan. Makassar and Balikpapan are the project that already gained success and Serang is new project which new products development will be proposed to launch. Based on profiling and product preference of customer in Balikpapan, Makassar, and prospectus of Serang markets, new products development that will be proposed are house type of 120/200 m² with price around Rp1.300.000.000 and house type of 71/120 m² with price around Rp800.000.000. The markets of Serang and Balikpapan have similarities in profiles as urban city so the new products development will adopt the succeed story of Balikpapan project.

1. Introduction

Nowadays, almost all company operations using information technology to manage all transaction data. Currently information technologies not only as a business support in business operation but also become one of strategic planning to gain business success. Data records which become huge should give added value in management decision in making knowledge-driven decisions to improve current business or develop new business. To capture the market in new business development the company must know what they know in this digital era.

A long with database development, a treasure is found in data tomb. Data become precious when knowledge is found. The process called Knowledge Data Discovery; it is a nontrivial process of identifying valid, novel, potentially useful and ultimately understandable pattern in data [1].
In new business development, marketing is one of organization which identifies unfulfilled needs and preferences, define and measures their interest, determines which target market the organization can best serve, decides on appropriate products, services and program to serve these markets, and calls upon everyone in the organization to think and serve the customer [2]. In order to develop new product the marketing function needs to be fulfilled wisely. Data transaction can be one of resources in gaining knowledge from pattern of customer data. Therefore, based on knowledge of data discovery, the new product can be proposed to fit with the market. In the processes which involve multi criteria decision to decide appropriate products, another technique is needed as decision support system. AHP technique decomposes a complicated problem into a multilevel hierarchical structure of objective, criteria and alternatives [3].

In property industry, a concept of city development is crucial. Creating the new product development that matches with customer preference becomes a big challenge since it related with big investments in a long term. In the competitive market, data mining will give an insight for management to help find the profile of customers and find their preference, and selecting the new product development appropriate with their preference.

2. Data Mining
   Data mining methodology used in this research is CRISP-DM that stands for CRoss-Industry Standard Process for data mining process was developed by Daimler Chrysler, SPSS and NCR. CRISP-DM methodology consists on a cycle that comprises six stages as shown in Figure 1 [4]:

   ![CRISP-DM Methodology](image)

   Figure 1. CRISP-DM Methodology (Chapman et al., 2000)

   1. Business understanding – First phase to understanding business objectives and requirement of business perspective to define data mining task in achieving the objectives.
   2. Data understanding – phase to collect data and understanding the data to identifying data quality and discover data insights.
   3. Data preparation – phase to prepare raw data set to final data set which clean and qualified to process.
   4. Modelling – phase to apply modelling techniques that appropriate to data mining task.
   5. Evaluation – phase to evaluate the model whether the model is properly achieves the business objective.
   6. Deployment – phase to deploy the result to another process.

3. AHP
   AHP is a multi-criteria decision making method developed by Saaty and is one of the best known which allows both qualitative and quantitative approaches to solve complex decision problems [3].
The method is modelling the complex problem into a hierarchical structure which consists of the goal, criteria and sub criteria. The subjective evaluations are converted into score on 1-9 scaling method and calculate the weight then we can rank the values.

4. Case studies
Case studies of this research are real estate projects in Serang, Makassar, and Balikpapan. Makassar and Balikpapan are the project that already gain succeeds and Serang is new project which new products development will be proposed to launch. Data used in this research is shown in Table 1 as follow:

| NO | YEAR TRANSACTION | DATASET 1 SERANG | DATA SET 2 MAKASSAR | DATA SET 3 BALIKPAPAN |
|----|------------------|------------------|---------------------|-----------------------|
| 1  | 2008             | 0                | 0                   | 25                    |
| 2  | 2009             | 80               | 0                   | 33                    |
| 3  | 2010             | 123              | 105                 | 88                    |
| 4  | 2011             | 85               | 203                 | 105                   |
| 5  | 2012             | 68               | 193                 | 60                    |
| 6  | 2013             | 76               | 159                 | 28                    |
| 7  | 2014             | 99               | 92                  | 8                     |
| 8  | 2015             | 40               | 153                 | 3                     |
| **Total Data** | **571**          | **905**          | **350**             |

To propose new product development in real estate project, the methodologies used in this research are described in Figure 2, as follows:

| Phase 1: | CRISP - Data Mining : Finding profile of customer in Serang Finding product preference in Balikpapan & Makassar |
|----------|----------------------------------------------------------|
| Phase 2: | Literature Review : Market analysis in market trend and need Criteria new product proposed |
| Phase 3: | AHP : Ranking new product proposed |

**Figure 2. Phases of proposed methodology**

4.1. Phase 1 – CRISP - Data Mining
Phase of CRISP – Data mining for this research as follow:
1. Business understanding, the data mining technique used to profiling the customer in Serang and finding product preferences in Makassar and Balikpapan.
2. Data understanding, to reach the goal the data used has been adjusted. The data field that not needed are eliminated
3. Data preparation, to get demography information such as province, age, gender we break the attribute of ID number (KTP) which has 16 digits as follow: ABCDEFGHIJKLMNOP.
   AB : Defines province where the citizen live
   CD : Defines city/kabupaten where the citizen live
   EF : Defines kecamatan where the citizen live
   GH : Defines birth date and show the gender too, if the citizen is female then the birth date will date will be added by 40.
   KL : Define year of citizen birth
   MNOP : Define the sequential number based on same birth date in area where the KTP issued.
4. Modelling, the data mining modelling is using Weka with data mining task are classification and clustering. The data modelling and analysis results are as shown in Table 2.

| Data Analysis          | Serang                   | Makassar                        | Balikpapan                        |
|------------------------|--------------------------|---------------------------------|-----------------------------------|
| Customer Profile       | Status:                  |                                 |                                   |
|                        | 74.44% married           | Makassar customer are originally from Makassar. | Urban city come from many regions in Indonesia |
|                        | 23.81% single            | Their ID presents their economic dweller, and they prefer to buy property in their home town. |                                   |
|                        | 1.75% widow              |                                 |                                   |
|                        | ID town variance from many regions in Indonesia |                                 |                                   |
| Product preference     | First buyer for single home with price range Rp160 – 180 mio with occupation Karyawan or better type price range Rp190 – 300 mio | Home with price range Rp 200 – 400 mio | First buyer: price around Rp500 mio LB less than 102 m (range 49 – 113m), |
|                        |                         |                                 |                                   |
|                        | Up grader for married, bigger house price range Rp 300 mio – Rp 700 mio |                                 | Upgrader: price around Rp800 mio – Rp6 bio |

5. Evaluation, the model are evaluated with business development team since data represented need to be validated, and data mining is an iterative process.
6. Deployment, the model are deployed to next process AHP.

4.2. Literature Review

Result of Indonesia Property Sentiment Survey H2-2015 by Rumah123.com is 65% people ages 21-30 are motivated to own their house [5]. The unique is they buy to invest and then upgrade the house. 29% of first buyer is young people and the up grader is 6% ages 31-40. The main indicator in this survey is income. People with income less than Rp100.000.000 in a year tend to find house with price less than Rp300.000.000. But people with income more than Rp1.000.000.000/year are often to looking for the property with that range price, because it’s easy to reinvest. Based on Cushman and Wakefield Indonesia in 2015, customers preference in middle of 2015 are house type 45/60 and 120/115 with price range Rp600.000.000 – Rp1.200.000.000 [6].

Based on customers profiles of Serang analysis, Serang customers have buying power of property with price range Rp259.325.000 and up to Rp737.361.000, the citizens are new urban from different province in Indonesia. Their age more than 29 years old, worked as PNS, worker in chemical industry at Cilegon, and international company in Dubai, Europe, and Asia. In Serang, there are few developers that provides suitable home for up grader.

4.3. AHP

With the consideration of customer profile and literature review new product development are defined and built as product selection matrix shown in Table 3.
Table 3. Product selection matrix

| Type   | Price Range: Rp300.000.000 | LB/LT: 27/60 m² | Design: Modern minimalis | Facilities: Cluster Gate |
|--------|----------------------------|----------------|--------------------------|--------------------------|
| Type 1 | Price Range: Rp400.000.000 | LB/LT: 36/72 m² | Design: Modern minimalis | Facilities: Cluster Gate |
| Type 2 | Price Range: Rp800.000.000 | LB/LT: 71/120 m²| Design: Modern Classic   | Facilities: Cluster Gate, neighbourhood CCTV |
| Type 3 | Price Range: Rp1.300.000.000 | LB/LT: 120/200 m²| Design: Modern Classic   | Facilities: Cluster Gate, neighbourhood CCTV |

In each group, with the business development member team, scoring the 1-9 scaling method, then count the weight.

The highest weight score is Type 4 with score 0.586 and for option the second type is Type 3 with score 0.224 as shown in Table 4. In Serang, properties available are small to medium type by local developers, there are opportunities to market bigger house and complete facilities.

Table 4. Pair comparison of houses type

| House types | T1 | T2 | T3 | T4 | WEIGHT |
|-------------|----|----|----|----|--------|
| Type 1 (T1) | 1  | 2  | 0.5| 0.2| 0.118 |
| Type 2 (T2) | 0.5| 1  | 0.2| 0.2| 0.073 |
| Type 3 (T3) | 2  | 5  | 1  | 0.2| 0.224 |
| Type 4 (T4) | 5  | 5  | 5  | 1  | 0.586 |
| CR           |    |    |    |    | 0.088 |

Price element is the element which has highest weight with score 0.548 and (LB/LT) has the second element with score 0.256 as shown in Table 5. For first buyer segment, price will be the important element for consideration and then the building area and land area (LB/LT). They will accept the detail product offered.

Table 5. Pair wise comparison Type 1

| Type 1 | Price T1 | LB/LT T1 | Design T1 | Facilities T1 | WEIGHT |
|--------|----------|----------|-----------|---------------|--------|
| Price T1 | 1       | 3        | 5         | 7             | 0.548  |
| LB/LT T1 | 1/3     | 1        | 3         | 5             | 0.256  |
| Design T1 | 1/5     | 1/3      | 1         | 5             | 0.145  |
| Facilities T1 | 1/7     | 1/5      | 1/5       | 1             | 0.052  |
| CR          |         |          |           |               | 0.093  |

Comparing criteria element in Type 2 shown in Table 6, which has more expensive than Type 2, the highest weight important element is Facilities with score 0.520 and the second important is Design with score 0.297. For first buyer who have more money or up grader they will start consider the facilities provided.

Table 6. Pair wise comparison Type 2

| Type 2 | Price T2 | LB/LT T2 | Design T2 | Facilities T2 | WEIGHT |
|--------|----------|----------|-----------|---------------|--------|
| Price T2 | 1       | 1/3      | 1/5       | 1/7           | 0.058  |
| LB/LT T2 | 3       | 1        | 1/3       | 1/5           | 0.124  |
| Design T2 | 5       | 3        | 1         | ½             | 0.297  |
| Facilities T2 | 7       | 5        | 2         | 1             | 0.520  |
| CR          |         |          |           |               | 0.026  |
Comparing criteria element in Type 3 shown in Table 7, which has more expensive than Type 1 and Type 2, the highest weight important element is Facilities with score 0.589 and the second important is Design with score 0.216. For market segment with high price, the facilities become high consideration.

**Table 7.** Pair wise comparison Type 3

| Type 3 | Price T3 | LB/LT T3 | Design T3 | Facilities T3 | WEIGHT |
|--------|----------|----------|-----------|---------------|--------|
| Price T3 | 1 | ½ | 1/3 | 1/5 | 0.080 |
| LB/LT T3 | 2 | 1 | 1/3 | 1/5 | 0.116 |
| Design T3 | 3 | 3 | 1 | 1/5 | 0.216 |
| Facilities T3 | 5 | 5 | 5 | 1 | 0.589 |
| CR=0.081 |

Comparing criteria element in Type 4 shown in Table 8, which have more expensive than Type 1 and Type 2 and Type 3. The highest weight important element is LB/LT with score 0.473 and the second important is Design with score 0.298. The consideration for highest price product will more detail, where markets want more in space, facilities and design.

**Table 8.** Pair wise comparison Type 4

| Type 4 | Price T4 | LB/LT T4 | Design T4 | Facilities T4 | WEIGHT |
|--------|----------|----------|-----------|---------------|--------|
| Price T4 | 1 | 1/5 | 1/3 | ½ | 0.087 |
| LB/LT T4 | 5 | 1 | 2 | 3 | 0.473 |
| Design T4 | 3 | ½ | 1 | 3 | 0.298 |
| Facilities T4 | 2 | 1/3 | 1/3 | 1 | 0.142 |
| CR=0.024 |

**Table 9.** Weight of selected product

| Weight of each group of Elements | Weight of each element | Total weight |
|----------------------------------|------------------------|--------------|
| **TYPE 4 0.586** | | |
| Price Range: Rp 1,300,000,000 | 0.087 | 0.050 |
| LB/LT: 120/200 m² | 0.473 | 0.277 |
| Design: Modern Classic | 0.298 | 0.175 |
| Facilities: Cluster Gate, neighbourhood CCTV | 0.142 | 0.083 |
| **TYPE 3 0.224** | | |
| Price Range: Rp 800,000,000 | 0.080 | 0.017 |
| LB/LT: 71/120 m² | 0.116 | 0.026 |
| Design: Modern Classic | 0.216 | 0.048 |
| Facilities: Cluster Gate, neighbourhood CCTV | 0.589 | 0.132 |
| **TYPE 1 0.118** | | |
| Price Range: Rp 300,000,000 | 0.548 | 0.065 |
| LB/LT: 37/60 m² | 0.256 | 0.030 |
| Design: Modern minimalist | 0.145 | 0.017 |
| Facilities: Cluster Gate | 0.052 | 0.006 |
| **TYPE 2 0.073** | | |
| Price Range: Rp 400,000,000 | 0.058 | 0.004 |
| LB/LT: 45/60 m² | 0.124 | 0.009 |
| Design: Modern minimalist | 0.297 | 0.022 |
| Facilities: Cluster Gate | 0.520 | 0.038 |

Finally, the overall priority scores are calculated as shown in Table 9. Type 4 is the proposed new product developments which have the highest weight and the second type option is Type 3. The opportunity of exclusive product with brand, facilities and design price range Rp800 mio –
Rp1.3 bio will be consider for new product development proposed in Serang.

5. Conclusion

The research shows that data mining process finds the profile of Serang customers and products preference in Balikpapan and Makassar project. Serang and Balikpapan has similarity as urban city, the market growth since there is opportunity in industry that attracted urban people as economic dweller to live in new city. Price range and product type of Makassar project has few variances customizes buying power of the market. Serang as new urban city need an exclusive product, with buying power up to Rp750.000.000, big house in type 3 and 4 can be proposed as new product development.

From experiment that has been done, data transaction contains valuable information that can be use in reselling, target marketing, and customer relationship management for future works.

However new product development is a complex process, since there many factor that not captured in data. Economic situation, government regulation, and customer behaviours in changed markets. Other technique is needed to complement the process.

References

[1] Fayyad, U.M., Smyth, P. and Uthurusamy, R. eds., 1996. Advances in knowledge discovery and data mining (Vol. 21). Menlo Park: AAAI press.
[2] Kotler, P., 2011. Marketing insights from A to Z: 80 concepts every manager needs to know. John Wiley & Sons.
[3] Saaty, T.L., 1988. What is the analytic hierarchy process? In Mathematical models for decision support (pp. 109-121). Springer, Berlin.
[4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinzart, T., Shearer, C. and Wirth, R., 2000. CRISP-DM 1.0 Step-by-step data mining guide.
[5] Kompas. (2015, November 3). Rumah Rp300 Jutaan Diincar Anak Muda Usia 21-30 Tahun. July 2016. http://properti.kompas.com/read/2015/11/03/150000021/ Rumah.Rp.300.Jutaan. Diincar.Anak.Muda.Usia.21-30.Tahun
[6] Peluang Properti. (2015, October 15). Rumah yang jadi incaran para konsumen. July 2016. http://www.peluangproperti.com/berita/kpr/2015-10/5898/rumah-yang-jadi-incaran-para-konsumen#sthash.qRUAbLaLB.dpuf
[7] Jain, S., Aalam, M.A. and Doja, M.N., 2010. K-means clustering using weka interface. In Proceedings of the 4th National Conference (pp. 25-26).
[8] Rahadi, R.A., Wiryono, S.K., Koesrindartoto, D.P. and Syamwil, I.B., 2012. Relationship between consumer preferences and value propositions: a study of residential product. Procedia-Social and Behavioral Sciences, 50, pp.865-874.
[9] Rahadi, R.A., Wiryono, S.K., Koesrindartoto, D.P. and Syamwil, I.B., 2013. Attributes influencing housing product value and price in Jakarta metropolitan region. Procedia-Social and Behavioral Sciences, 101, pp.368-378.
[10] Singh Swati, Dubey Gaurav, “Finding interest people in purchasing real estate by using data mining techniques”. International Journal of Recent Technology and Engineering(IJRTE) ISSN:2277-3878, Volume X, Issue X.
[11] Ho, H.P., Chang, C.T. and Ku, C.Y., 2015. House selection via the internet by considering homebuyers’ risk attitudes with S-shaped utility functions. European Journal of Operational Research, 241(1), pp.188-201.
[12] Sener, H.Y., 2014. Determining New Markets Using Analytic Hierarchy Process: Case Study in Gürul Porcelain. International Journal of Marketing Studies, 6(5), p.149.
[13] Padhy, N., Mishra, D. and Panigrahi, R., 2012. The survey of data mining applications and feature scope. arXiv preprint arXiv:1211.5723.
[14] Chimbalkar, A., Patil, S., Jain, M., Kate, G. and Dwivedi, P., 2014. Finding n-Most Demanding Products for Real-Estate Industry. International Journal of Computer Applications, 93(12).