

# Association rules of Prognostic Variables for Survival in a Randomized Comparison of Treatments for Prostatic Cancer

Iing Lukman<sup>1\*)</sup>, Emy Khikmawati<sup>2</sup>

<sup>1</sup>School of Management, Faculty of Economics, Universitas Malahayati

<sup>2</sup>School of Industrial Engineering, Faculty of Engineering, Universitas Malahayati

<sup>1,2</sup>Jln. Pramuka 27 Bandarlampung, Indonesia

email: <sup>1</sup>deiingofthelukman@gmail.com, <sup>2</sup>emy\_khikmawati@yahoo.com

Copyright ©2019, Politeknik Harapan Bersama, Tegal

**Abstract** – In this paper, the data was analyzed by data mining techniques of association rules. The data for 506 patients consist of an identification number, stage of tumour, a code for the treatment to which the patient was assigned, the date of randomization, the total months of follow-up since randomization, an indicator for the survival status or cause of death, and the values of twelve pretreatment covariates. The goal of an analysis should be to compare the treatments with respect to survival of the patients. Since this was a randomized study it would ordinarily not be necessary to adjust for the values of the pretreatment covariates. However, in such studies it is advisable to examine the prognostic significance of the covariates and to confirm that they are balanced across treatment groups. In addition, the analyst should look for important treatment-covariates interactions which might lead to the definition of subsets of patients in which treatment differences were significantly more marked or even reversed.

**Abstrak** – Dalam artikel ini, data dianalisis dengan teknik data mining aturas asosiasi. Data untuk 506 pasien terdiri dari nomor identifikasi, stadium tumor, kode untuk perawatan yang diberikan kepada pasien, tanggal pengacakan, total bulan follow-up sejak pengacakan, indikator untuk status bertahan hidup atau penyebab kematian, dan nilai-nilai dari dua belas pretreatment covariates. Tujuan dari analisis ini adalah membandingkan treatments dengan kelangsungan hidup pasien, dikarenakan ini adalah penelitian acak, maka tidak menyesuaikan nilai covariates pretreatment. Namun pada studi tersebut disarankan untuk memeriksa signifikansi dari covariates prognostic dan mengkonfirmasi bahwa hasilnya seimbang diantara treatment groups. Selain itu analisis ini mencari interaksi treatment-covariates yang penting yang dapat mengarah pada definisi himpunan bagian pasien dimana perbedaan pengobatan secara signifikan lebih ditandai atau baha terbalik.

**Kata Kunci** – data mining, prognostic, cancer, randomized clinical trial.

---

\*) **Corresponding Author:** (Iing Lukman)  
Email: deiingofthelukman@gmail.com

## I. INTRODUCTION

A prognosis is the doctor's best estimate of how cancer will affect someone. Many factors can affect a person's prognosis. In general, survival is used to estimate the percentage of people with cancer who will live at least a certain amount of time (such as 1, 3, 5, or 10 years) after their diagnosis [1], [2]. The independent prognostic factors affecting survival were assessed in 240 men undergoing treatment for metastatic prostate cancer as part of a randomized clinical trial comparing the gonadotropin releasing hormone analogue Zoladex (goserelin acetate implant) with castration [3].

In multivariate analysis, the most highly significant predictors were the presence or absence of bone pain, serum testosterone levels, serum alkaline phosphate levels, and performance status. Patients with all forms favorable for survival had a 2-year survival rate of 84% as compared with only 8% for patients with none of the few factors favorable for survival, no other factors were significant [3].

## II. RESEARCH METHODOLOGY

The material was data from Andrews and Herzberg (1985) [4]. The data was the comparison of treatment for prostatic cancer. The layout of data was as follows in Table 1.

The variables involved are as follows: Pat. No. is patient number. Mos FU is complete months of follow-up. SBP is systolic blood pressure. DBP is diastolic blood pressure. EKG is electrocardiogram. HG is serum haemoglobin. SZ is size of primary tumour. SG is combined index of tumour stage and histologic grade. AP is serum prostatic acid phosphatase in King-Amstrong units. BM is bone metastases. Furthermore, the next page told the variables also.

The mining process was conducted by rapidminer version 5.2 software [5]. The aim of applying Association Rules was to detect relationships or associations between specific values of categorical variables in the large data sets [6], [7]. This technique allows analysts and researchers to uncover hidden patterns in large data sets.

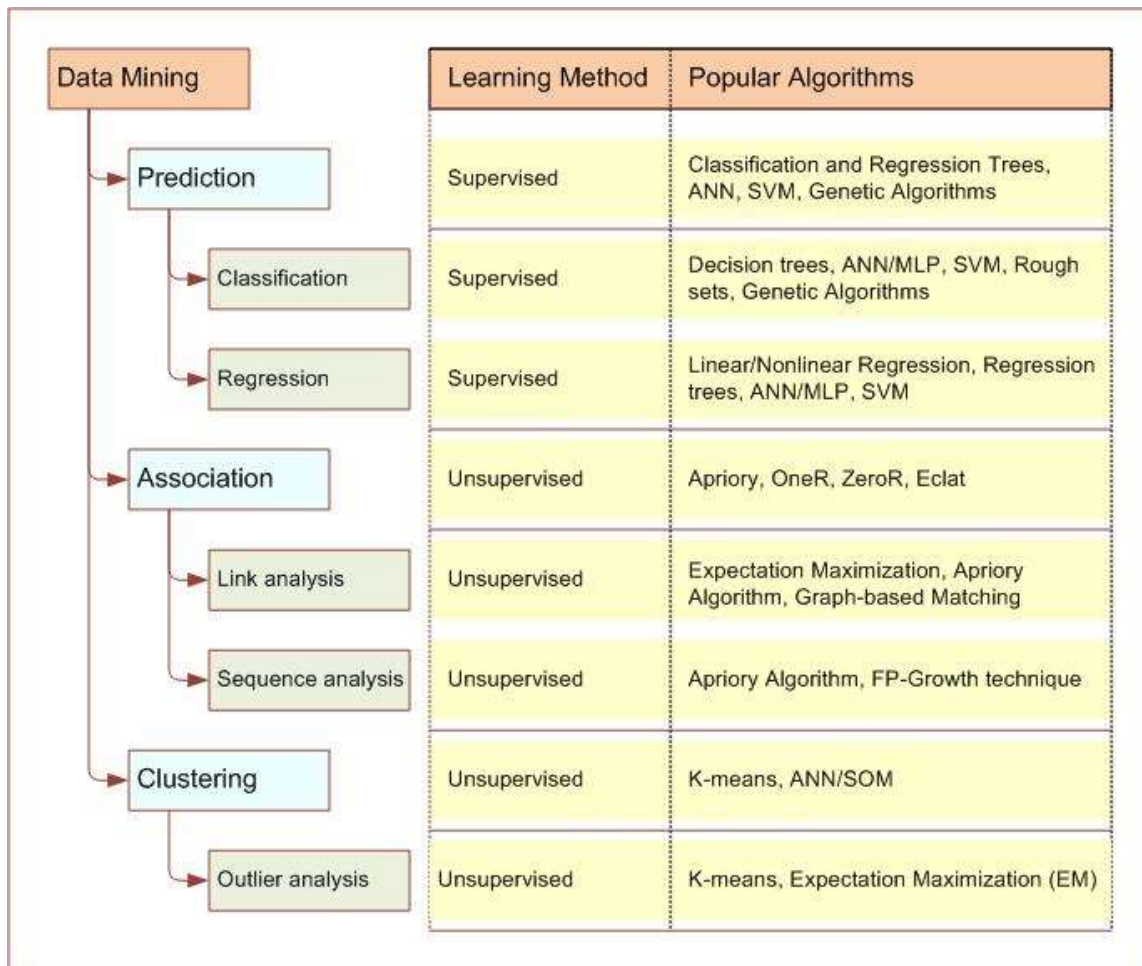


Figure. 1 Taxonomy of Data Mining Tasks

TABLE I  
COMPARISON TREATMENTS FOR PROSTATIC CANCER

| Pat. No. | Stage | Rx | Date On study | Mos FU | Surv Stat | Pretreatment covariates |    |    |    |       |       |       |     |    |     |     |    |
|----------|-------|----|---------------|--------|-----------|-------------------------|----|----|----|-------|-------|-------|-----|----|-----|-----|----|
|          |       |    |               |        |           | Age years               | Wt | PF | HX | S B P | D B P | E K G | HG  | SZ | S G | AP  | BM |
| 1        | 3     | 2  | 81067         | 72     | 0         | 75                      | 76 | 0  | 0  | 15    | 9     | 4     | 138 | 2  | 8   | 3   | 0  |
| .        | .     | .  | .             | .      | .         | .                       | .  | .  | .  | .     | .     | .     | .   | .  | .   | .   | .  |
| .        | .     | .  | .             | .      | .         | .                       | .  | .  | .  | .     | .     | .     | .   | .  | .   | .   | .  |
| .        | .     | .  | .             | .      | .         | .                       | .  | .  | .  | .     | .     | .     | .   | .  | .   | .   | .  |
| 506      | 4     | 3  | 52169         | 4      | 1         | 74                      | 97 | 1  | 0  | 15    | 6     | 0     | 212 | 33 | -   | 222 | 1  |

### III. RESULTS AND DISCUSSION

The 506 cases of data sets were shown in the Figure 3. From the 18 attribute (variables) inputs we obtained 221 Association Rules.

From more than 405960 association rules generated by the data and later the minimum criterion imposed on 0.8 for its support and confidence then come out 221 results as can be seen on Figure 3 where the highest support and highest confidence lied on cases 212 to 221. If the premises was each (Stage), (Rx), (Pat no), (Date on-study), (AP) then the conclusion were Age,yrs, Wt, Mos FU. It meant that if the premises each was the stage of prostatic cancer, the Rx (the treatment), Patient Number, Date on-study, and AP (Serum Prostatic Acid) then the conclusion were age of the patient, Weight index of the patients, and the number of complete months of follow-up. It meant that its premises and its conclusion created association rules which its support and

confidence of 0.955 or 95.5% each. It also if the premises was (Stage, Rx) then the conclusion was Age,yrs, Wt, Mos FU. It meant that the Stage and the treatment of the cancer were associated with the age, weight index, and the number of complete months of follow up with its support and confidence of .955 or 95.5%. If the premises was Stage only, then the conclusion was (Pat no, Age,yrs, Wt, and Mos FU). It meant that the stage was associated with Patient Number, Age of patients, weight index of the patients and the number of complete months of follow-up. If the premises were Patient Number then the conclusion was Stage, and age, weight index, and the number of complete months of follow-up

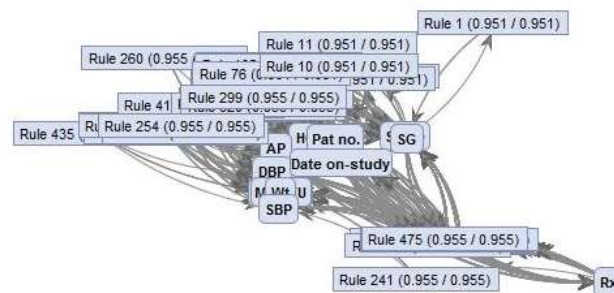


Figure. 2 Association Rules architecture for prognostic variables data set

| No  | Premises                              | Conclusion                   | Support | Confid. | LaPl. | Gain   | p-s | Lift | Consu. |
|-----|---------------------------------------|------------------------------|---------|---------|-------|--------|-----|------|--------|
| 1   | Stage                                 | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 2   | Rx                                    | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 3   | Pat no.                               | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 4   | Date on-study                         | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 5   | AP                                    | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 6   | Stage                                 | Rx, SG, SZ                   | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 7   | Rx                                    | Stage, SG, SZ                | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 8   | Stage, Rx                             | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 9   | Stage                                 | Pat no., SG, SZ              | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 10  | Pat no.                               | Stage, SG, SZ                | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 11  | Stage, Pat no.                        | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 12  | Stage                                 | Date on-study, SG, SZ        | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 13  | Date on-study                         | Stage, SG, SZ                | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 14  | Stage, Date on-study                  | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 15  | Stage                                 | AP, SG, SZ                   | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 16  | AP                                    | Stage, SG, SZ                | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 17  | Stage, AP                             | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 18  | Rx                                    | Pat no., SG, SZ              | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 19  | Pat no.                               | Rx, SG, SZ                   | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 20  | Rx, Pat no.                           | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 201 | Stage, Pat no., AP                    | Rx, Date on-study, SG, SZ    | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 202 | Rx, Pat no., AP                       | Stage, Date on-study, SG, SZ | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 203 | Stage, Rx, Pat no., AP                | Date on-study, SG, SZ        | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 204 | Date on-study, AP                     | Stage, Rx, Pat no., SG, SZ   | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 205 | Stage, Date on-study, AP              | Rx, Pat no., SG, SZ          | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 206 | Rx, Date on-study, AP                 | Stage, Pat no., SG, SZ       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 207 | Stage, Rx, Date on-study, AP          | Pat no., SG, SZ              | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 208 | Pat no., Date on-study, AP            | Stage, Rx, SG, SZ            | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 209 | Stage, Pat no., Date on-study, AP     | Rx, SG, SZ                   | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 210 | Rx, Pat no., Date on-study, AP        | Stage, SG, SZ                | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 211 | Stage, Rx, Pat no., Date on-study, AP | SG, SZ                       | 0.951   | 0.951   | 0.975 | -1.045 | 0   | 1    | 1      |
| 212 | Stage                                 | Age,yrs, Wt, Mos FU          | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 213 | Rx                                    | Age,yrs, Wt, Mos FU          | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 214 | Pat no.                               | Age,yrs, Wt, Mos FU          | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 215 | Date on-study                         | Age,yrs, Wt, Mos FU          | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 216 | AP                                    | Age,yrs, Wt, Mos FU          | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 217 | Stage                                 | Rx, Age,yrs, Wt, Mos FU      | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 218 | Rx                                    | Stage, Age,yrs, Wt, Mos FU   | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 219 | Stage, Rx                             | Age,yrs, Wt, Mos FU          | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 220 | Stage                                 | Pat no., Age,yrs, Wt, Mos FU | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |
| 221 | Pat no.                               | Stage, Age,yrs, Wt, Mos FU   | 0.955   | 0.955   | 0.977 | -1.045 | 0   | 1    | 1      |

Figure. 3 The rules matching between its premises and its conclusion



## AssociationRules

### Association Rules

```
[Stage] --> [SG, SZ] (confidence: 0.951)
[Rx] --> [SG, SZ] (confidence: 0.951)
[Pat no.] --> [SG, SZ] (confidence: 0.951)
[Date on-study] --> [SG, SZ] (confidence: 0.951)
[AP] --> [SG, SZ] (confidence: 0.951)
[Stage] --> [Rx, SG, SZ] (confidence: 0.951)
[Rx] --> [Stage, SG, SZ] (confidence: 0.951)
[Stage, Rx] --> [SG, SZ] (confidence: 0.951)
[Stage] --> [Pat no., SG, SZ] (confidence: 0.951)
[Pat no.] --> [Stage, SG, SZ] (confidence: 0.951)
[Stage, Pat no.] --> [SG, SZ] (confidence: 0.951)
[Stage] --> [Date on-study, SG, SZ] (confidence: 0.951)
[Date on-study] --> [Stage, SG, SZ] (confidence: 0.951)
[Stage, Date on-study] --> [SG, SZ] (confidence: 0.951)
[Stage] --> [AP, SG, SZ] (confidence: 0.951)
[AP] --> [Stage, SG, SZ] (confidence: 0.951)
[Stage, AP] --> [SG, SZ] (confidence: 0.951)
[Rx] --> [Pat no., SG, SZ] (confidence: 0.951)
[Pat no.] --> [Rx, SG, SZ] (confidence: 0.951)
[Rx, Pat no.] --> [SG, SZ] (confidence: 0.951)
[Rx] --> [Date on-study, SG, SZ] (confidence: 0.951)
[Date on-study] --> [Rx, SG, SZ] (confidence: 0.951)
[Rx, Date on-study] --> [SG, SZ] (confidence: 0.951)
[Rx] --> [AP, SG, SZ] (confidence: 0.951)
[AP] --> [Rx, SG, SZ] (confidence: 0.951)
[Rx, AP] --> [SG, SZ] (confidence: 0.951)
```

(a)

```
[Stage, AP] --> [Date on-study, SG, SZ] (confidence: 0.951)
[Date on-study, AP] --> [Stage, SG, SZ] (confidence: 0.951)
[Stage, Date on-study, AP] --> [SG, SZ] (confidence: 0.951)
[Rx] --> [Pat no., Date on-study, SG, SZ] (confidence: 0.951)
[Pat no.] --> [Rx, Date on-study, SG, SZ] (confidence: 0.951)
[Rx, Pat no.] --> [Date on-study, SG, SZ] (confidence: 0.951)
[Date on-study] --> [Rx, Pat no., SG, SZ] (confidence: 0.951)
[Rx, Date on-study] --> [Pat no., SG, SZ] (confidence: 0.951)
[Pat no., Date on-study] --> [Rx, SG, SZ] (confidence: 0.951)
[Rx, Pat no., Date on-study] --> [SG, SZ] (confidence: 0.951)
[Rx] --> [Pat no., AP, SG, SZ] (confidence: 0.951)
[Pat no.] --> [Rx, AP, SG, SZ] (confidence: 0.951)
[Rx, Pat no.] --> [AP, SG, SZ] (confidence: 0.951)
[AP] --> [Rx, Pat no., SG, SZ] (confidence: 0.951)
[Rx, AP] --> [Pat no., SG, SZ] (confidence: 0.951)
[Pat no., AP] --> [Rx, SG, SZ] (confidence: 0.951)
[Rx, Pat no., AP] --> [SG, SZ] (confidence: 0.951)
[Rx] --> [Date on-study, AP, SG, SZ] (confidence: 0.951)
[Date on-study] --> [Rx, AP, SG, SZ] (confidence: 0.951)
[Rx, Date on-study] --> [AP, SG, SZ] (confidence: 0.951)
[AP] --> [Rx, Date on-study, SG, SZ] (confidence: 0.951)
[Rx, AP] --> [Date on-study, SG, SZ] (confidence: 0.951)
[Date on-study, AP] --> [Rx, SG, SZ] (confidence: 0.951)
[Rx, Date on-study, AP] --> [SG, SZ] (confidence: 0.951)
[Pat no.] --> [Date on-study, AP, SG, SZ] (confidence: 0.951)
[Date on-study] --> [Pat no., AP, SG, SZ] (confidence: 0.951)
[Pat no., Date on-study] --> [AP, SG, SZ] (confidence: 0.951)
... 405960 other rules ...
```

(b)

Figure. 4 (a),(b) The Summary of Association rules which its confidence of 0.

TABLE II  
SUMMARY OF ASSOCIATION RULES WHICH ITS SUPPORT AND CONFIDENCE OF 0.955

| Premises           | Conclusion                  | Support | Confidence |
|--------------------|-----------------------------|---------|------------|
| Stage, Pat, No, AP | Age,yrs, Wt, Mos FU         | 0.955   | 0.955      |
| Rx                 | Age,yrs, Wt, Mos FU         | 0.955   | 0.955      |
| Pat no             | Age,yrs, Wt, Mos FU         | 0.955   | 0.955      |
| Date on-study      | Age,yrs, Wt, Mos FU         | 0.955   | 0.955      |
| AP                 | Age,yrs, Wt, Mos FU         | 0.955   | 0.955      |
| Stage              | Rx, Age,yrs, Wt, Mos FU     | 0.955   | 0.955      |
| Rx                 | Stage, Age,yrs, Wt, Mos FU  | 0.955   | 0.955      |
| Stage, Rx          | Age,yrs, Wt, Mos FU         | 0.955   | 0.955      |
| Stage              | Pat no, Age,yrs, Wt, Mos FU | 0.955   | 0.955      |
| Pat no             | Stage, Age,yrs, Wt, Mos FU  | 0.955   | 0.955      |

## IV. CONCLUSION

The output of association rules with minimum criterion 0.8 resulted in 221 rules with 10 association rules could be considered as the most probable accepted association between premises and its conclusion due to its support and its confidence was the biggest one of 95.5% each. However, from the ten biggest score of support and confidence, for every determined premises the conclusion was always contained age, weight index and the number of complete months of follow up of the patients.

## ACKNOWLEDGMENT

Herewith, we are highly appreciated to Universitas Malahayati for providing fund of this research project.

## REFERENCES

- [1] D. P. Byar and S. B. Green, "The choice of treatment for cancer patients based on covariate information," *Bull. Cancer*, vol. 67, no. 4, pp. 477–90, 1980.
- [2] D. P. Byar and D. K. Corle, "Selecting optimal treatment in clinical trials using covariate information," *J. Chronic Dis.*, vol. 30, no. 7, pp. 445–459, Jul. 1977.
- [3] G. W. Chodak, "Independent Prognostic Factors in Patients With Metastatic (Stage D2) Prostate Cancer," *JAMA J. Am. Med. Assoc.*, vol. 265, no. 5, p. 618, Feb. 1991.
- [4] A. David F and H. Agnes M, *Data: a collection of problems from many fields for the student and research worker*. Springer Science & Business Media, 2012.
- [5] A. Jovic, K. Brkic, and N. Bogunovic, "An overview of free software tools for general data mining," *2014 37th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2014 - Proc.*, no. May, pp. 1112–1117, 2014.
- [6] A. M. Zamani and B. Amaliah, "Implementasi Algoritma Genetika pada Struktur Backpropagation Neural Network untuk Klasifikasi Kanker Payudara," *J. Tek. POMITS*, vol. 1, no. 1, pp. 1–6, 2012.
- [7] N. Robert, E. John, and G. Miner, *Handbook of statistical analysis and data mining applications*. Academic Press, 2009.