MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 12, No. 1 (2020), pp 71-74 DOI : 10.18860/mat.v12i1.8932
Received : January25th 2020; **Accepted** : March 9th 2020 ; **Avalaible Online** : April 5th 2020

Pendekatan Box - Whisker Plot dan Regresi Liniear untuk Prediksi User Upgrade pada Start Up AyoCPNS

Puspa Miladin Nuraida Safitri A. Basid, Fajar Rohman Hariri

***Abstract*— Many start-ups in Indonesia began to become an important part of an economy. In the implementation of startups is still developing and requires a lot of funding for operational. So these start-up companies need to think about steps to survive and develop. Besides through investors, they also carry out promotional steps to attract users. To do the promotion itself, this startup needs to learn how the impact of the promotion that has been done. It's enough or not, it can be seen from the number of users per day. Besides predicting users who will upgrade the premium account also need to be done. This research has produced an equation to make these predictions using linear regression and Box-Whisker Plots.**

***Index Terms*— Box-Whisker Plots, linear regression, start-up**

***Abstrak*–- Munculnya banyak start up di indonesia mulai menjadi bagian penting dari perekonomian sebuah negara. Dalam pelaksanaanya perusahaan rintisan seperti ini masih berkembang dan membutuhkan banyak pendanaan untuk operasionalnya. Sehingga para perusahaan rintisan ini perlu memikirkan langkah untuk berkembang lagi. Selain melalui investor, merka juga melakukan langkah promosi untuk menarik user. Untuk melakukan promosi itu sendiri, perusahaan rintisan ini perlu memperhatikan bagaimana dampak dari promosi yang telah dilakukan. Sudah cukup atau belum, hal tersebut dapat di lihat dari jumlah user perharinya. Selain itu memprediksi user yang akan upgrade accont premium juga perlu dilakukan. Penelitian ini telah menghasilkan persamaan untuk melakukan prediksi tersebut dengan menggunakan metode regresi linier dan Box-Whisker Plot.**

***Kata Kunci*— Box-Whisker Plot, regresi linier, , start up**

I. INTRODUCTION

Dewasa ini, banyak bermunculan *start-up* baru di kalangan penduduk indonesia. Apabila membahas tentang start-up, mak yang akan muncul di benak kita adalah beberapa perusahaan yang sudah dikenal seperti Gojek, Tokopedia, Tiket.com, Traveloka dan lain sebagainya. Start-up sendiri adalah perusahaan rintisan yang meluncurkan produk inovatif [1]. Di indonesia sendiri mulai banyak bermunculan start-upa berbasis teknologi digital. Mereka muncul dengan berbagai inovasi yng dibutuhkan oleh masyarakat. Mulai dari start-up aplikasi pelayanan cuci mobil hingga start-up yang menawarkan bimbingan belajar secara online. Salah satunya adalah website ayocpns yang menawarkan produk layanan bimbingan belajar secara online bagi masyarakat yang akan mengikuti ujian cpns.

Perusahaan rintisan seperti ini secara tidak disadari mulai menjadi bagian penting dari perekonomian sebuah negara [2]. Namun, dalam pelaksanaanya perusahaan rintisan seperti ini masih berkembang dan membutuhkan banyak pendanaan untuk operasionalnya. Beberapa perusahaan rintisan memanfaatkan investor untuk melakukan pendanaan. Namun bila ditelisik lebih dalam, untuk menjaring investor hal yang perlu diperhatikan adaah seberapa banyak user yang tertarik menggunakan layanan dari perusahaan ini. Ini juga merupakan salah satu problema yang dihadapi oleh perusahaan rintisan selain pendanaan [3]. Peneliti menganggap pemasaran produk atau layanan masih memegang kunci utama dalam menggaet pengguna. Dalam melakukan pemasaran produk atau layanan sebuah start-up, penulis menganggap perlunya sebuah prediksi jumlah pengguna. Hal ini salah satunya ditujukan untuk menjadi tolak ukur seberapa efektifkah pemasaran yang dilakukan sehingga pemasaran lebih optimal dan tidak mengalami pemborosan.

Salah satu perusahan rintisan yang sedang naik daun saat ini adalah website ayocpns.com. Jika diliat dari awal rilis (November, 2019) hingga kini (Februari 2020) memiliki pengguna sejumlah 287.000, hanya 60-75% user yang mau meng-upgrade accountnya menjadi account premium berbayar. Sehingga perlu menentukan seberapa banyak pemasaran yang harus dilakukan. Peneliti menganggap perlu adanya cara untuk menentukan prediksi jumlah user untuk upgrade ke account premium sebagai acuan pelaksaanan pemasaran.

Pada penelitian ini, penulis melakukan prediksi jumlah user upgrade pada perusahaan rintisan ayocpns

Manuscript received March 05, 2020.This work was supported in part by Informatics Engineering Department of Maulana Malik Ibrahim Islamic State University.

Puspa Miladin Nuraida Safitri.A.B Author is with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University , Malang, Indonesia; email puspa.miladin@uin-malang.ac.id

Fajar Rohman Hariri Author, was Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University. (e-mail: dosendeso@gmail.com).

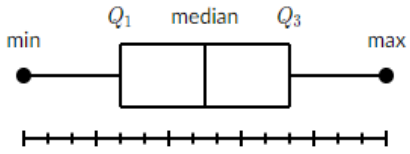
menggunakan regresi liniear. Namun sebelum itu perlu diterapkan pendekatan Box - Whisker Plot untuk *data cleaning* agar memperoleh hasik prediksi yang maksimal. Hal ini diharapkan dapat membantu perusahaan rintisan ini untuk merencanakan teknik pemasaran ang lebih optimal.

II. LITERATURE REVIEW

Prediksi adalah estimasi atau perkiraan mengenai suatu keadaan yang akan datang. Salah satu keuntungan melakukan prediksi adalah untuk merencanakan langkah selanjutnya atau kebutuhan yang akan dibutuhkan mendatang. Dalam melakukan prediksi berkaitan langsung dengan "apa yang diminta" dan "berapa banyak dan kapan harus disediakan" [4]. Salah satu teknik untuk melakukan prediksi adalah menggunakan regresi liniear. Metode ini merupakan bentuk hubungan variabel bebas X atau variabel Y sebagai faktor berpangkat satu. Pada regresi linier sederhana, satu variabel independen dan satu variabel dependen terlibat [5]. Variabel yang akan diprediksi disebut sebagai variabel respon, sedangkan variabel yang menerangkan disebut variabel bebas [6]. Variabel respon Y dan variabel bebas X1, X2, X3, …,Xn memiliki hubungan dapat dilihat pada persamaan model regresi liniear adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 X_I + \varepsilon_i \text{ , i = 1,2,.....,n} \quad (1) \quad [7]$$

Namun sebelum melakukan prediksi dengan regresi linear, peneliti menganggap perlunya melakukan analisis data dengan menggunakan metode Box - Whisker Plot. Tujuanya dalah untuk menegtahui pemusatan dan sebaran data dari nilai tengah dan nilai outliernya. Peneliti menganggap dengan adanya outlier pada data yang akan diolah, menjadi penyebab model regresi kurang baik. Maka penulis berendapat perlu dilakukan pendeteksian outlier pada data.



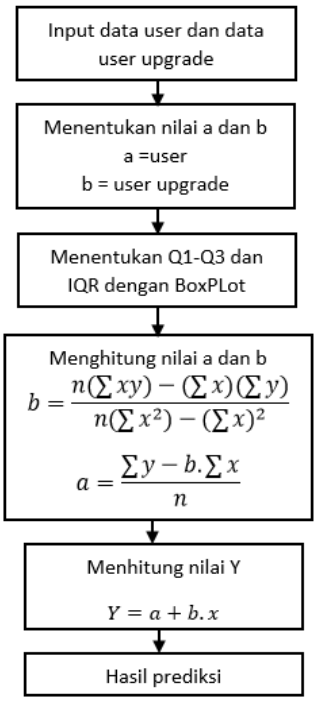
Gambar. 1. Model Box - Whisker Plot [8]

Garis vertical bagian kiri box menunjukan kuartil pertama (Q1), sementara kanan menunjukan kuartil ketiga (Q3). Sedangkan box ialah bidang yang merepresentasikan interquartile range (IQR), atau bagian tengah. Lebar bagian box ini ditentukan oleh IQR. IQR merupakan representasi dari ukuran sebaran data. Apabila ukuran box semakin lebar, menunjukkan data sebaran data semakin lebar. Sedangkan garis horizontal yang melewati bidang box merepresentasikan median dari data. Kemudian garis yang mempertinggi bagian box disebut whiskers. Whiskers inilah yang menunjukan nilai lebih rendah dan lebih tinggi dari data IQR. Untuk panjang garis Whisker bagian atas ini adalah ≤ Q3 + (1.5 x IQR). Panjang garis Whisker bagian kiri adalah ≥ Q1 – (1,5 x IQR). Pada akhirnya nilai yang berada di kanan atau kiri whisker disebut dengan nilai outlier.

III. METHOD

Dalam pelaksanaannya, diagram blok pada gambar 2 menunjukan bagaimana alur kerja dari penelitian ini. Dimulai dari inputan data yang merupaka data user pengguna website ayocpns. Kemudian menentukan nilai a yang merupakan data user non upgrade dan b adalah data user upgrade account premium. Data yang di ambil adalah data perhari mulai november 2019 hingga februari 2020.

Kemudian langkah selanjutnya adalah melakukan *data cleaning* untuk menentukan nilai ekstrem/outlier dengan metode box-whisker plot. Setelah data didapat maka ditentukan nilai a dan b untuk mendapatkan persamaan regresi linear. Setelah itu barulah prediksi dapat ditentukan dengan menggunaan persamn regresi linear yang sudah ada.



Gambar. 2. Alur proes

IV. EXPERIMENTAL RESULT

A. *Box - Whisker Plot*

Data mentah user yang diperoleh dari website ayocpns berjumlah 120, beberapa terdapat data yang ekstrem. Dikatakan data ekstrem adalah ketika pada periode hari tersebut tidak ada data yang masuk, tidak ada user yang upgrade account premium, dan jumlah data yang terlalu besar disbanding data lainya. Hal ini d anggap akan mengakibatkan hasil prediksi yang kurang baik, sehingga dilakukan proses *data cleaning* untuk mengilangkan data ekstrem tersebut. Tujuanya adalah guna merubah data ke menjadi suatu format yang nantinya akan membuat proses lebih efektif untuk mendapatkan nilai prediksi yang lebih akurat. Selain itu hal ini juga akan mengurangi waktu perhitungan untuk

data berskala besar. Tabel 1 menunjukan sampel data awal sebelum dilakukan proses Box-Whisper Plot.

Table 1. Sampel data user

Waktu	User Register	User Upgrade Premium
01/11/2019	38	37
04/11/2019	2	1
07/11/2019	1	1
10/11/2019	1	1
11/11/2019	11	7
12/11/2019	1	1
18/11/2019	39	1
19/11/2019	17	3
20/11/2019	6	2
21/11/2019	6	1
22/11/2019	5	3
23/11/2019	905	119
24/11/2019	924	221
25/11/2019	663	178
26/11/2019	1028	151
27/11/2019	482	117
28/11/2019	309	78
29/11/2019	242	71
30/11/2019	304	61
01/12/2019	180	53
02/12/2019	655	93
….	….	…..
04/12/2019	577	99
05/12/2019	407	89
06/12/2019	443	76
07/12/2019	627	77
08/12/2019	1257	165
09/12/2019	944	164
10/12/2019	710	160
….	….	…..
26/01/2020	2998	406
27/01/2020	6928	952
28/01/2020	6451	1149
29/01/2020	6660	830
30/01/2020	10351	1962
31/01/2020	9861	2708
01/02/2020	7008	1737
02/02/2020	7664	1715
03/02/2020	9442	2050
04/02/2020	9250	2506
05/02/2020	12537	2259
06/02/2020	12975	1986
07/02/2020	10265	1808

Dilakukan perhitungan dengan metode Box-Whisper Plot seperti di bawah ini:

$$Q3 + (1{,}5\,.IQR) < outlier \leq Q3 + (3\,.IQR) \quad (2)$$

$$Q1 - (1{,}5\,.IQR) < outlier \leq Q1 - (3\,.IQR) \quad (3)$$

Diperoleh nilai:
Q1= 740,75
Q3= 3185,75
IQR = 2445
Batas bawah = Q1-(1,5 x IQR) = -2926,75
Batas atas = Q3+(1,5 x IQR)= 6853,25

Sehingga data yang diluar range batas atas dan batas bawah akan dihapus. Tabel 2 menunjukan hasil dari data yang telah di cleaning dengan Box-Whisper Plot. Data awal yang berjumlah 120, setelah praproses berubah menjadi 85. Angka yang jumlahnya diluar range menghilang.

Table 2. Sampel data user setelah praproses

User Register	User Upgrade Premium
6	2
6	1
5	3
905	119
924	221
663	178
1028	151
482	117
309	78
242	71
304	61
180	53
655	93
704	117
577	99
407	89
443	76
627	77
1257	165
944	164
710	160
736	199

B. Regresi Liniear

Metode ini memiliki tujuan untuk menguji sejauh mana hubungan antara variable penyebab (x) dan variable respon (y). Persamaan dari metode ini digambarkan seperti berikut:

$$Y = a + b\text{X} \quad (4)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (5)$$

$$a = \frac{\sum y - b.\sum x}{n} \quad (6)$$

y = variable respon
x = variabel bebas
a = konstanta
b = besaran respon yang disebabkan oleh variable bebas.
n = jumlah data
∑y = jumlah data y
∑xy = jumlah data xy

Setelah data diolah dengan persamaan di atas, maka nilai:

a = -39,9811382
b = 0,206609087

sehingga persamaan untuk menentukan prediksi user upgrade account premium pada website ayopns adalah

$$y = -39{,}9811382 + 0{,}206609087\,.X \quad (7)$$

X pada persamaan di atas merupak jumlah awal user yang tidak melakukan upgrade account.

Table 3. Data Prediksi dan MAE

User Register	User Upgrade Premium	REGRESI (Prediksi)	REAL	MAE	MSE
704	117	105,4717	117	11,52834	132,9026
577	99	79,23231	99	19,76769	390,7618
407	89	44,10876	89	44,89124	2015,223
443	76	51,54669	76	24,45331	597,9645
627	77	89,56276	77	12,56276	157,8229
1257	165	219,7265	165	54,72648	2994,988
944	164	155,0578	164	8,94216	79,96222
710	160	106,7113	160	53,28869	2839,684
736	199	112,0832	199	86,91685	7554,539
790	152	123,24	152	28,75996	827,1353
943	132	154,8512	132	22,85123	522,1788
714	126	107,5378	126	18,46225	340,8547
892	81	144,3142	81	63,31417	4008,684

1241	229	216,4207	229	12,57926	158,2378
1356	282	240,1808	282	41,81922	1748,847
1226	290	213,3216	290	76,6784	5879,577
1055	251	177,9914	251	73,00855	5330,249
891	191	144,1076	191	46,89244	2198,901
755	169	116,0087	169	52,99128	2808,075
1411	157	251,5443	157	94,54428	8938,622
2473	578	470,9631	578	107,0369	11456,89
1614	509	293,4859	509	215,5141	46446,31
1395	80	248,2385	80	168,2385	28304,21
1531	683	276,3374	683	406,6626	165374,5
1564	470	283,1555	470	186,8445	34910,88
1598	452	290,1802	452	161,8198	26185,65
1716	449	314,5601	449	134,4399	18074,1
1658	830	302,5767	830	527,4233	278175,3
1963	1076	365,5925	1076	710,4075	504678,8
1184	300	204,644	300	95,35598	9092,763
1566	376	283,5687	376	92,43131	8543,547
1505	391	270,9655	391	120,0345	14408,27
1180	475	203,8176	475	271,1824	73539,9
1278	333	224,0653	333	108,9347	11866,77
1415	491	252,3707	491	238,6293	56943,93
1590	565	288,5273	565	276,4727	76437,15
2400	500	455,8807	500	44,11933	1946,515
1732	431	317,8658	431	113,1342	12799,35
1618	442	294,3124	442	147,6876	21811,64
1554	445	281,0894	445	163,9106	26866,69
1427	314	254,85	314	59,14997	3498,719
2050	301	383,5675	301	82,56749	6817,391
2594	740	495,9628	740	244,0372	59554,14
2027	490	378,8155	490	111,1845	12362
2198	535	414,1456	535	120,8544	14605,78
2032	484	379,8485	484	104,1515	10847,53
1604	365	291,4198	365	73,58016	5414,04
1519	338	273,8581	338	64,14193	4114,188
1979	425	368,8982	425	56,10175	3147,407
3222	538	625,7133	538	87,71334	7693,63
3608	606	705,4644	606	99,46445	9893,177
3202	617	621,5812	617	4,581159	20,98702
3137	590	608,1516	590	18,15157	329,4794
2607	590	498,6488	590	91,35125	8345,05
2998	406	579,4329	406	173,4329	30078,97
4206	723	829,0167	723	106,0167	11239,54
4139	654	815,1739	654	161,1739	25977,02
3988	559	783,9759	559	224,9759	50614,16
4363	586	861,4543	586	275,4543	75875,08
3528	508	688,9357	508	180,9357	32737,74
3033	462	586,6642	462	124,6642	15541,17
2674	489	512,4916	489	23,49156	551,8534
2170	377	408,3606	377	31,36058	983,4861
			Rerata	102,365	22808,52
			MAE		
			MSE		

Dari hasil prediksi yang diperoleh, kemudan di tentukan nilai MAE dan MSE nya. MAE dan MSE merepresentasikan rata – rata kegagalan (eror) absolut dari hasil peramalan dengan nilai sebenarnya.

V. CONCLUSION

Berdasarkan penelitian yang telah dilakukan, di dapatkan persamaan untuk prediksi user upgrade premium account yaitu y=-39,9811382+0,206609087 .X. Hasil prediksi dari persamaan ini dapat digunakan sebagai acuan untuk melakukan kegiatan promosi guna meningkatkan user yang melakukan upgrade account pada website ayocpns. Dari persamaan tersebut juga setelah dilakukan ujicoba di dapatkan nilai MAE sebesar 102,365 dan MSE 22808,52 sehinngga dianggap cukup akurat.

REFERENCES

[1] E. Klotins, M. Unterkalmsteiner, and T. Gorschek, "Software-Intensive Product Engineering in Start-Ups: A Taxonomy," *IEEE Softw.*, vol. 35, no. 4, pp. 44–52, 2018, doi: 10.1109/MS.2018.2801548.

[2] S. Srinivasan, I. Barchas, M. Gorenberg, and E. Simoudis, "Venture Capital: Fueling the Innovation Economy," *Computer (Long. Beach. Calif).*, vol. 47, no. 8, pp. 40–47, 2014, doi: 10.1109/MC.2014.230.

[3] C. Giardino, S. S. Bajwa, X. Wang, and P. Abrahamsson, "Key Challenges in Early-Stage Software Startups," *Springer Int. Publ.*, vol. 212, pp. 52–63, 2015, doi: 10.1007/978-3-319-18612-2.

[4] V. Gaspersz, *Production Planning And Inventory Control*. Jakarta: PT. Gramedia Pustaka Utama, 2005.

[5] M. S. Acharya, A. Armaan, and A. S. Antony, "A comparison of regression models for prediction of graduate admissions," *ICCIDS 2019 - 2nd Int. Conf. Comput. Intell. Data Sci. Proc.*, pp. 1–5, 2019, doi: 10.1109/ICCIDS.2019.8862140.

[6] D. N. Gujarati, *Dasar-Dasar Ekonometrika*, 1st ed. Jakarta: Penerbit Erlangga, 2007.

[7] N. Draper and H. Smith, *Applied Regresssion Analysis*, 2nd ed. Jakarta: Gramedia Pustaka Utama, 1992.

[8] K. Academy, "Box-Plot Review." [Online]. Available: https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/box-whisker-plots/a/box-plot-review.