

# Penerapan Data Mining Untuk Analisis Pengaruh Nilai Imla dan Nahwu Terhadap Nilai Bahasa Arab Di Pesantren

Yahdinil Firdha Nadiroh

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Serang Raya

Email : yahdinilll14@gmail.com

## ABSTRAKSI

Bahasa arab merupakan Bahasa yang sering digunakan sehari-hari oleh santri di pesantren. Bahkan wajib digunakan untuk berkomunikasi. Hal itu mengharuskan santri untuk belajar lebih giat lagi serta lebih banyak menghafal kosa kata Bahasa arab. Dengan adanya penelitian ini bertujuan untuk menerapkan data mining menggunakan metode regresi linier berganda guna menganalisis pengaruh nilai Imla dan Nahwu terhadap nilai Bahasa Arab pada santriwati di pesantren. Data penelitian ini menggunakan sampel dari nilai santriwati yang berupa nilai Imla, Nahwu dan Bahasa Arab. Metode ini diterapkan untuk mengidentifikasi hubungan serta pengaruh ke-tiga nilai tersebut apakah dapat mempengaruhi Tingkat keberhasilan dalam Pelajaran Bahasa Arab. Dengan adanya data mining dapat mengetahui pengaruh serta hubungan diantara ke-tiga Pelajaran tersebut. Dan dengan adanya penelitian ini diharapkan memberikan kontribusi positif bagi pesantren dan Lembaga Pendidikan sejenis dalam Upaya peningkatan mutu pembelajaran Bahasa Arab. Selain itu, metode data mining yang diterapkan dapat dijadikan referensi untuk penelitian selanjutnya dalam menggali informasi lebih lanjut tentang faktor-faktor penentu keberhasilan belajar Bahasa Arab di lingkungan pesantren.

**Kata Kunci:** data mining, regresi linier, orange, pengaruh nilai, nahwu, imla, bahasa arab

## ABSTRACT

Students in Islamic boarding schools often use Arabic as their daily language. It is even mandatory to use it to communicate. This requires students to study harder and memorize more Arabic vocabulary. This study aims to apply data mining using multiple linear regression methods to analyze the effect of Imla and Nahwu values on Arabic language values in female students in Islamic boarding schools. This study used samples of female students' values in Imla, Nahwu, and Arabic. This method is applied to identify the relationship and influence of the three values and whether they can affect the level of success in Arabic language lessons. With data mining, the influence and relationship between the three lessons can be determined. And with this study, it is expected to provide a positive contribution to Islamic boarding schools and similar educational institutions in efforts to improve the quality of Arabic language learning. In addition, the data mining method applied can be used as a reference for further research exploring further information about the determining factors for the success of learning Arabic in Islamic boarding schools.

**Keywords:** data mining, linear regression, orange, value influence, nahwu, dictum, arabic

## Penulis Korespondensi

Yahdinil Firdha Nadiroh

Tanggal Submit : 21/01/2024

Tanggal Diterima : 11/07/2024

Tanggal Terbit : 26/07/2024

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## I. PENDAHULUAN

Data mining diartikan merupakan sekumpulan proses yang berguna mengeksplorasi dan mencari nilai yang berupa informasi serta relasi-relasi kompleks yang selama ini tersimpan dari suatu basis data. Dengan meakukan penggalian pola informasi terhadap data yang berguna sebagai memanipulasi data menjadi sebuah informasi baru serta lebih bermanfaat yang didapatkan melalui cara mengekstrasi juga mengetahui

beberapa pola yang berharga atau menarik yang didapatkan dari basis data [1]. Pondok Pesantren Modern merupakan Lembaga pendidikan Agama Islam dengan sistem asrama di mana Kyai sebagai figur, masjid sebagai tempat ibadahnya, dan pembelajaran agama islam serta Bahasa arab di bawah bimbingan Kiyai yang diikuti santri sebagai kegiatan utamanya.

Pondok pesantren merupakan salah satu lembaga pendidikan berbasis islam yang ada di Indonesia, yang

di dalamnya mengajarkan berbagai macam pelajaran keagamaan mengenai islam dan sebagai salah satu lembaga yang berperan banyak dalam pendidikan moral dan akhlak yang mulia bagi para santri didalamnya [2].

Pondok pesantren mengajarkan disiplin tinggi kepada para santri. Para santri dilatih untuk disiplin, taat beragama, patuh kepada orang tua, dan taat terhadap peraturan yang berlaku di pesantren. Pendidikan Bahasa Arab di lingkungan pesantren memiliki peran yang sangat penting dalam membentuk generasi muslim yang mampu memahami dan menguasai Bahasa Arab sebagai sarana untuk memahami Al-Quran serta ajaran Islam. Pencapaian tingkat penguasaan Bahasa Arab di Tingkat Madrasah Tsanawiyah (MTs) di pesantren menjadi hal krusial dalam menentukan kualitas Pendidikan Lembaga tersebut. Harapan akhirnya adalah semua santri dapat mengasah serta mempelajari Pelajaran Bahasa arab dengan maksimal. Namun, setiap santri memiliki Tingkat keterampilan yang berbeda sehingga memerlukan bimbingan yang berbeda pula. Oleh karena itu, penelitian ini memfokuskan pada penerapan Data Mining metode Regresi Linier Berganda untuk menganalisis pengaruh nilai Imla dan Nahwu terhadap nilai Bahasa Arab pada santriwati MTs di pesantren. Regresi linier berganda adalah hubungan secara linier antara dua atau lebih variable independen dengan variable (Y).

Tujuannya dilakukan penelitian ini adalah memberi kontribusi pada pesantren dengan mengintegrasikan pendekatan data mining untuk menganalisis faktor-faktor pengaruh dalam pembelajaran Bahasa Arab di pesantren. Membantu pesantren dalam mengidentifikasi area pembelajaran yang memerlukan perhatian khusus, sehingga dapat merancang strategi pengajaran yang lebih efisien.

## II. METODE PENELITIAN

Penelitian ini menggunakan beberapa tahapan adapun tahapan seperti yang diperlihatkan pada Gambar 1.

### 2.1 Sumber Data

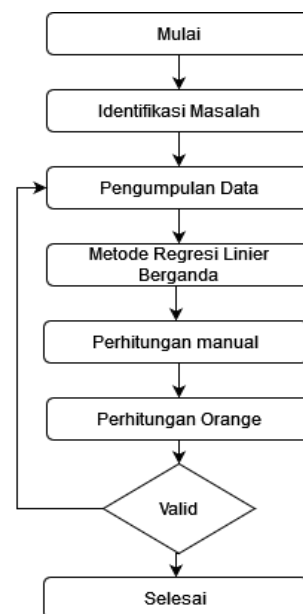
Metode penelitian ini menggunakan Metode Regresi Linier berganda dan Penelitian ini dilakukan di Pondok Pesantren Darunnajah 3 yang merupakan pesantren modern yang terletak di Jalan Palka Km.07 Kecamatan Pabuaran Kabupaten Serang Provinsi Banten, menempati areal seluas 3,5 hektar dan merupakan salah satu cabang Pondok Pesantren Darunnajah Jakarta atau lebih di kenal Darunnajah Pusat. Pondok pesantren Darunnajah 3 didirikan pada tanggal 26 Januari 1992 oleh Bapak KH. Abdul Manaf Mukhayyar (alm). Sampel data yang digunakan pada penelitian ini adalah nilai ujian akhir semester 1 2023 kelas 2-3 MTs santriwati Darunnajah 3. Diketahui bahwa variabel X1 adalah nilai Imla, dan variabel X2 merupakan nilai Nahwu sedangkan variabel Y adalah nilai Bahasa Arab. Adapun datanya adalah seperti pada Tabel 1.

Tabel 1. Sample Data Penelitian

No	X <sub>1</sub>	X <sub>2</sub>	Y	X <sub>1</sub> <sup>2</sup>	X <sub>2</sub> <sup>2</sup>	Y <sup>2</sup>	X <sub>1</sub> X <sub>2</sub>	X <sub>1</sub> Y	X <sub>2</sub> Y
1	87	87	92	7569	7569	8464	7569	8004	8004
2	91	81	89	8281	6561	7921	7371	8099	7209
3	76	60	59	5776	3600	3481	4560	4484	3540
4	91	81	77	8281	6561	5929	7371	7007	6237
5	78	69	73	6084	4761	5329	5382	5694	5037
6	75	82	86	5625	6724	7396	6150	6450	7052
7	71	79	74	5041	6241	5476	5609	5254	5846
8	76	65	61	5776	4225	3721	4940	4636	3965
9	75	52	71	5625	2704	5041	3900	5325	3692
10	89	69	59	7921	4761	3481	6141	5251	4071
11	90	73	85	8100	5329	7225	6570	7650	6205
12	90	78	65	8100	6084	4225	7020	5850	5070
13	78	67	65	6084	4489	4225	5226	5070	4355
14	78	63	73	6084	3969	5329	4914	5694	4599
15	70	59	72	4900	3481	5184	4130	5040	4248
16	87	80	81	7569	6400	6561	6960	7047	6480
17	93	86	89	8649	7396	7921	7998	8277	7654
18	88	89	89	7744	7921	7921	7832	7832	7921
19	71	81	78	5041	6561	6084	5751	5538	6318
20	82	83	70	6724	6889	4900	6806	5740	5810
21	84	72	81	7056	5184	6561	6048	6804	5832
22	83	85	64	6889	7225	4096	7055	5312	5440
23	75	70	62	5625	4900	3844	5250	4650	4340
24	92	67	90	8464	4489	8100	6164	8280	6030
25	63	60	70	3969	3600	4900	3780	4410	4200
26	65	58	70	4225	3364	4900	3770	4550	4060
27	77	59	70	5929	3481	4900	4543	5390	4130
28	70	76	69	4900	5776	4761	5320	4830	5244
29	71	61	69	5041	3721	4761	4331	4899	4209
30	61	50	66	3721	2500	4356	3050	4026	3300
31	.....	.....	.....	.....	.....	.....	.....	.....	.....
150	64	67	62	4096	4489	3844	4288	3968	4154
151	48	58	61	2304	3364	3721	2784	2928	3538
152	44	53	58	1936	2809	3364	2332	2552	3074
153	42	50	51	1764	2500	2601	2100	2142	2550
154	39	48	48	1521	2304	2304	1872	1872	2304
155	35	62	59	1225	3844	3481	2170	2065	3658
156	44	38	47	1936	1444	2209	1672	2068	1786
	11523	10342	10727	884603	721224	760055	785405	807751	719911

### 2.2 Data Mining

Data mining merupakan proses menemukan informasi dari suatu data yang tersimpan dalam suatu database atau datasheet. Pembuatan model dilakukan dengan proses menggunakan algoritma atau rumus tertentu. Proses data mining menggunakan berbagai teknik seperti Teknik dalam proses statistic, matematika, dan machine learning yang digunakan dalam melakukan identifikasi dan mengelolah berbagai data menjadi informasi yang bermanfaat [3].



Gambar 1. Tahapan Penelitian

### 2.3 Regresi Linier Berganda

Analisis regresi linear berganda adalah teknik regresi yang memiliki banyak variabel bebas. Satu diantara keunggulan analisis regresi linear berganda mampu menduga keadaan di masa depan melalui pengukuran beberapa variabel bebas (X) dengan variabel tidak bebas (Y) [4].

Metode regresi linier berganda adalah suatu Teknik statistika yang digunakan untuk memodelkan hubungan antara dua atau lebih variabel independent dengan suatu variabel dependen. Regresi linier berganda memperluas konsep regresi linier sederhana, yang hanya melibatkan satu variabel independent dan satu variabel dependen. Tujuan dari regresi linier berganda adalah untuk mengetahui masing-masing variabel apakah memiliki hubungan positif atau negatif. Rumus regresi linier berganda dapat dirumuskan. Apabila terdapat dua variabel bebas yaitu X1 dan X2 maka bentuk persamaan regresinya seperti pada persamaan (1).

$$Y = a + b_1X_1 + b_2X_2 \quad (1)$$

Keterangan :

Y : Variabel terikat (nilai yang diprediksikan)

X<sub>1</sub>, X<sub>2</sub> : Variabel bebas

a : Konstanta (nilai Y' apabila semua X bernilai 0)

b : Koefisiensi regresi (nilai peningkatan ataupun penurunan)

Berikut adalah kondisi koefisien regresi yang mungkin terjadi, b<sub>1</sub> dan b<sub>2</sub> memiliki suatu nilai :

1. Nilai 0, hal tersebut terjadi apabila Y tidak dipengaruhi variabel X<sub>1</sub> dan X<sub>2</sub>.
2. Bernilai negative (-), hal itu disebabkan hubungan terbalik antara variabel Y dengan variabel X<sub>1</sub> dan X<sub>2</sub>.
3. Bernilai positif (+), hal itu disebabkan hubungan searah antara variabel Y dengan variabel X<sub>1</sub> dan X<sub>2</sub>.

### 2.4 Nahwu

Nahwu adalah ilmu tentang atau pokok, yang bisa diketahui dengannya akhir suatu kata baik secara i'rab atau bina. Ilmu nahwu adalah dalil-dalil yang memberitahu kepada kita bagaimana seharusnya keadaan akhir kata-kata itu setelah tersusun dalam kalimat, atau ilmu yang membahas kata-kata Arab dari i'rab dan bina' [5].

### 2.5 Imla

Metode Imlâ' disebut juga metode dikte atau metode menulis di mana guru mengucapkan materi pelajaran dan siswa disuruh menulisnya dibuku tulis. Imlâ' atau dictation, yaitu tahapan untuk mengukur kemampuan peserta didik untuk menuliskan apa yang didengarnya [6].

## IV. HASIL DAN PEMBAHASAN

Dari data tabel 2 sudah ditentukan nilai Imla menjadi variabel X<sub>1</sub> dan nilai Nahwu menjadi variabel X<sub>2</sub>. Sedangkan nilai Bahasa Arab menjadi variabel Y, variabel yang terkait (variabel dependent) atau di sebut dengan nilai yang akan diprediksi. Setelah menentukan variabelnya maka dibuat tabel tambahan seperti X<sub>1</sub><sup>2</sup>, X<sub>2</sub><sup>2</sup>, Y<sup>2</sup>, X<sub>1</sub>X<sub>2</sub><sup>2</sup>, X<sub>1</sub>Y, X<sub>2</sub>Y. Sebelum menentukan nilai b<sub>1</sub>, b<sub>2</sub>, a dan r<sup>2</sup> kita tentukan nilai a, b<sub>1</sub>, dan b<sub>2</sub> terlebih dahulu kemudian mencari nilai  $\sum X_1^2$ ,  $\sum X_2^2$ ,  $\sum Y^2$ ,  $\sum X_1Y$ ,  $\sum X_2Y$ , dan  $\sum X_1X_2$  yang diperlihatkan pada Tabel 2.

Tabel 2. Hasil Perhitungan Manual

$\sum X_1^2$	33452,17
$\sum X_2^2$	35602,44
$\sum Y^2$	22436,22
$\sum X_1Y$	15397,01
$\sum X_2Y$	8765,91
$\sum X_1X_2$	21489,19

Nilai yang telah didapatkan tersebut digunakan untuk menentukan nilai dari a, b<sub>1</sub> dan b<sub>2</sub> yang dapat dilihat pada penyelesaian matematika berikut :

#### 1. Nilai b<sub>1</sub>

$$b_1 = \frac{\{(\sum X_2^2 \times \sum X_1Y) - (\sum X_1Y \times \sum X_1X_2)\}}{\{(\sum X_1^2 \times \sum X_2^2) - (\sum X_1X_2)^2\}}$$

$$b_1 = \frac{\{(35602,44 \times 15397,01) - (15397,01 \times 21489,19)\}}{\{(33452,17 \times 35602,44) - (21489,19)^2\}}$$

$$b_1 = \frac{\{(548171124,7044) - (330869273,3219)\}}{\{(1190978875,2948) - (461785286,8561)\}}$$

$$b_1 = \frac{\{(217301851,3825)\}}{\{(729193588,4387)\}}$$

$$b_1 = 0,29$$

#### 2. Nilai b<sub>2</sub>

$$b_2 = \frac{\{(\sum X_1^2 \times \sum X_2Y) - (\sum X_1Y \times \sum X_1X_2)\}}{\{(\sum X_1^2 \times \sum X_2^2) - (\sum X_1X_2)^2\}}$$

$$b_2 = \frac{\{(33452,17 \times 8765,91) - (15397,01 \times 21489,19)\}}{\{(33452,17 \times 35602,44) - (21489,19)^2\}}$$

$$b_2 = \frac{\{(-37630561,7972)\}}{\{(729193588,4387)\}}$$

$$b_2 = -0,05$$

#### 3. Nilai a

$$a = \frac{(\sum Y) - (b_1 \times \sum X_1) - (b_2 \times \sum X_2)}{n}$$

$$a = \frac{(10727) - (0,29 \times 11523) - (-0,05 \times 10342)}{156}$$

$$a = \frac{7902,43}{156}$$

$$a = 50,65$$

#### 4. Nilai Koefisien Determinasi (r<sup>2</sup>)

$$r^2 = \frac{(b_1 \times \sum x_1 y) + (b_2 \times \sum x_2 y)}{\sum y^2}$$

$$r^2 = \frac{(0,29 \times 15397,01) + (-0,05 \times 8765,91)}{22436,22}$$

$$r^2 = \frac{4026,8374}{22436,22}$$

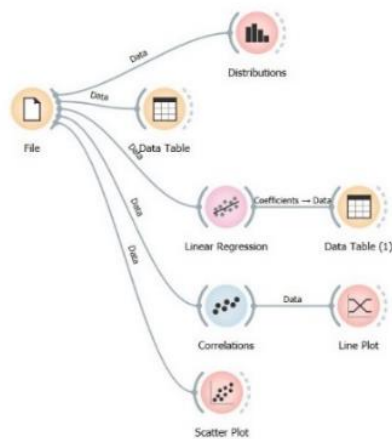
$$r^2 = 0,17$$

Jadi sekarang telah didapat nilai a, b1, b2 serta nilai koefisien determinasi ( $r^2$ ) sebagai berikut :

Tabel 3. Hasil Perhitungan Manual

$b_1$	0,29
$b_2$	-0,05
a	50,65
$r^2$	0,17

### 5. Hasil Perhitungan Orange

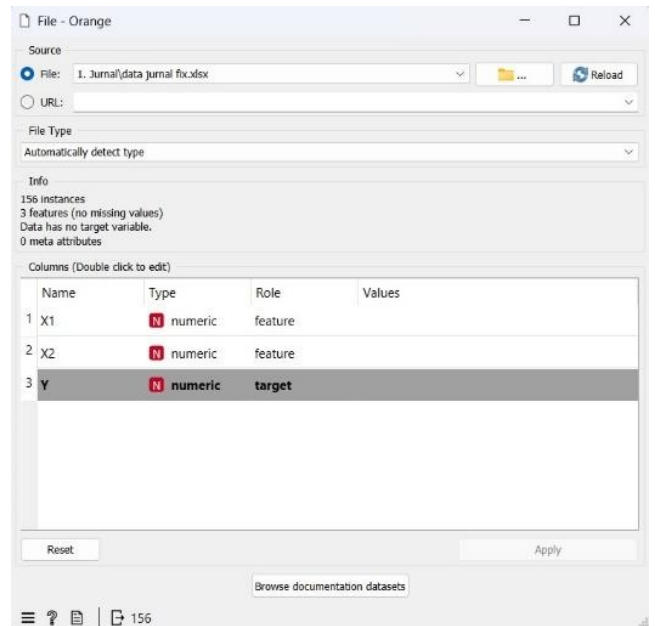


Gambar 2. Tampilan Orange

Dapat dilihat dari Gambar 2 merupakan tampilan dari aplikasi orange, terdapat file, distribution, data tabel, linear regression, correlations, line plot dan scatter plot. Berikut adalah hasil perhitungan dari aplikasi Orange :

#### 1) File

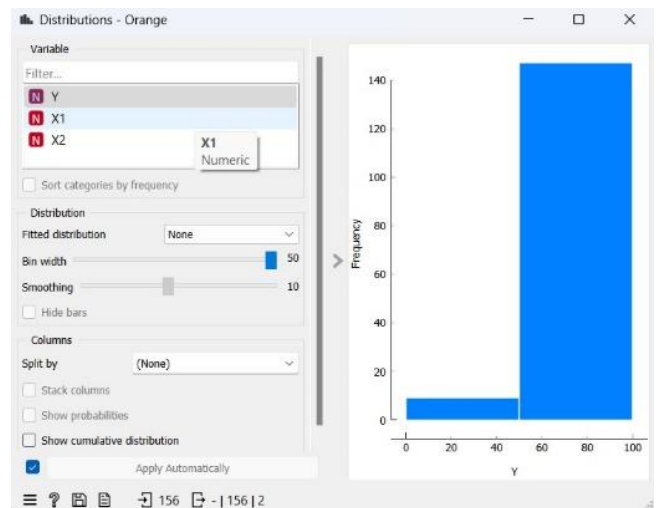
File ini berisi data nilai santri pada tabel 2 sebelumnya pindahkan data ke Microsoft excel lalu masukan data ke file orange lalu setelah data dimasukan terdapat tampilan pada gambar ke 2. Ubah Y menjadi target.



Gambar 2. File

#### 2) Distribution

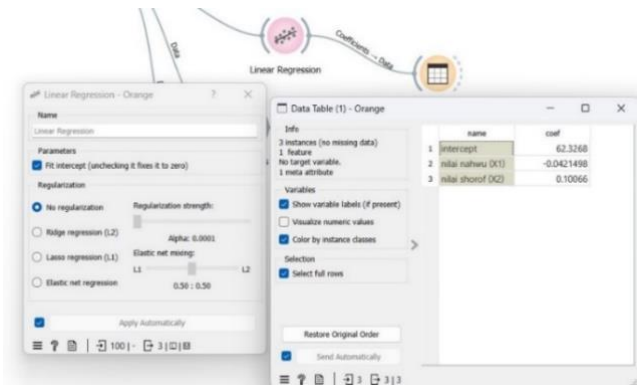
Distributions menampilkan distribusi nilai atribut diskrit atau kontinu. Jika data berisi variabel kelas, distribusi dapat dikondisikan pada kelas tersebut. Grafik menunjukkan berapa kali (misalnya, berapa kali) setiap nilai atribut muncul dalam data. Jika data berisi variabel kelas, distribusi kelas untuk setiap nilai atribut akan ditampilkan (seperti pada cuplikan di bawah) [7].



Gambar 3. Distribution

#### 3) Linier Regression

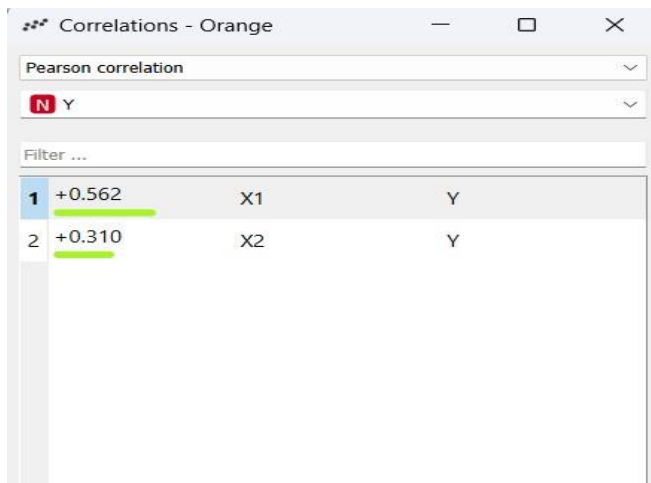
Linier regression mengidentifikasi hubungan antara variabel prediktor dan variabel respon Y.



Gambar 4. Linier regression

#### 4) Correlation

Correlation menunjukkan hubungan antara variabel. Variabel  $X_1$  terhadap variabel  $Y$  dan variabel  $X_2$  terhadap variabel  $Y$ . pada gambar 5 jelas bahwa masing masing variabel prediktor mempunyai hubungan terhadap variabel  $Y$ .



Gambar 5. Correlation

#### 5) Line Plot

Line plot digunakan untuk memvisualisasikan hubungan antara variabel prediktor dengan variabel respon dalam bentuk garis yang menghubungkan titik pada sumbu  $X_1$ ,  $X_2$  dan  $Y$ .

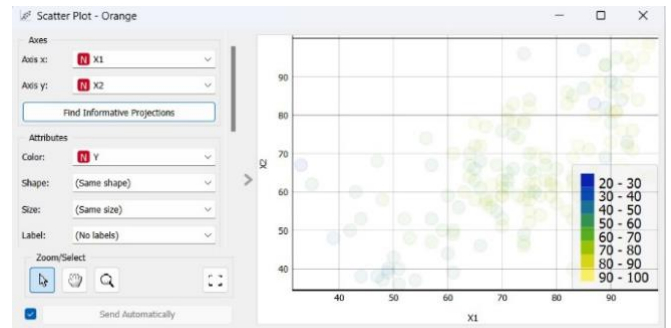


Gambar 6. Line Plot

#### 6) Scatter Plot

Scatter plot juga dikenal sebagai diagram sebar atau grafik sebar, adalah teknik visualisasi data dua

dimensi yang menampilkan hubungan antara dua variabel numerik. Grafik ini terdiri dari titik-titik yang masing-masing mewakili titik observasi atau data, diplot pada bidang kartesius dengan sumbu  $x$  dan sumbu  $y$ . Dengan memplot titik-titik ini dan memeriksa distribusinya, kita dapat melihat sifat dan kekuatan hubungan antar variabel. [8]



Gambar 7. Scatter Plot

Dari gambar 7 dapat disimpulkan bahwa hubungan nilai  $X_1$ ,  $X_2$  dan  $Y$  memiliki hubungan kolerasi yang positif karena titik data cenderung bergerak ke atas dari kiti ke kanan.

#### V. KESIMPULAN

Data mining dapat digunakan dalam membantu proses analisis pengaruh nilai Imla dan Nahwu terhadap nilai Bahasa Arab. Penelitian ini mengambil sampel nilai dari 156 Santri Tingkat Madrasah Tsanawiyah (MTs) di Pesantren Al-Manshur Darunnajah 3. Diketahui bahwa nilai Imla dan Nahwu saling berpengaruh terhadap nilai Bahasa Arab sehingga dapat di simpulkan bahwa nilai Imla dan Nahwu memiliki pengaruh terhadap nilai Bahasa Arab di pesantren. Hasil ini memberikan pengetahuan kepada santri untuk meningkatkan kualitas pembelajaran. Dan bagi pendidik untuk meningkatkan kualitas pengajarannya. Pada analisis pengaruh nilai Imla dan Nahwu terhadap nilai Bahasa Arab ini di buktikan melalui perhitungan manual dan perhitungan menggunakan aplikasi Orange bahwasannya nilai Imla dan Nahwu sangat berpengaruh dan saling berkaitan dengan nilai Bahasa Arab. Di perhitungan melalui aplikasi Orange pada gambar 6 correlation di situ jelas nilai correlation dari nilai Imla dan Bahasa Arab saling berpengaruh begitu pun dengan nilai Nahwu terhadap nilai Bahasa Arab saling berpengaruh.

#### DAFTAR PUSTAKA

- [1] Ditto 2019
- [2] Riskal Fitri & Syarifuddin Ondeng, 2022; Pesantren Di Indonesia : Lembaga Pembentukan Karakter.
- [3] Arhami & Nasir, 2020; Jollyta et al., 2020.
- [4] Sembiring, R. K., 1995, Analisis Regresi. Bandung : Penerbit ITB.

- [5] Ana Wahyuning Sari (2017), Analisis kesulitan pembelajaran Nahwu pada siswa kelas VIII MTs Al-Irsyad Gajah Demak Tahun Ajaran 2015/2016.
- [6] H.Guntur Cahaya Kesuma & Asti Fauziah (2016). PENERAPAN METODE IMLÂ' UNTUK MENINGKATKAN KETERAMPILAN MENULIS BAHASA ARAB PESERTA DIDIK KELAS X SMA IT PONDOK PESANTREN AL-MUJTAMA'AL-ISLAMI KARANG ANYAR LAMPUNG SELATAN.
- [7] [https://orange3-readthedocs.io.translate.google/projects/orange-visual-programming/en/latest/widgets/visualize/distributions.html?\\_x\\_tr\\_sl=en&\\_x\\_tr\\_tl=id&\\_x\\_tr\\_hl=id&\\_x\\_tr\\_pto=tc](https://orange3-readthedocs.io.translate.google/projects/orange-visual-programming/en/latest/widgets/visualize/distributions.html?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc)
- [8] [https://www-polymersearch-com.translate.google/glossary/scatter-plot?\\_x\\_tr\\_sl=en&\\_x\\_tr\\_tl=id&\\_x\\_tr\\_hl=id&\\_x\\_tr\\_pto=tc](https://www-polymersearch-com.translate.google/glossary/scatter-plot?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc)