# DATA SCIENCE: PENDEKATAN DAN LANGKAH PRAKTIS DENGAN EXCEL

**Ismail Setiawan[1], Aisyah Mutia Dawis[2]**

[1,2]Program Studi Sistem dan Teknologi Informasi Fakultas Sains Dan Teknologi,
Universitas 'Aisyiyah Surakarta, Surakarta, Indonesia 57146
Tlp . 0271 711270
Email: ismail@aiska-university.ac.id, aisyahmd@aiska-university.ac.id

***ABSTRACT***

*The steps in conducting data science activities consist of several stages, namely problem identification, understanding the current business, data collection, data processing, and making decisions based on insights. Researchers who engage in data science activities are often referred to as data scientists. In their process, data scientists use applications to facilitate their data science activities. One application that can be used by data scientists is Excel. Excel has features that can handle a certain amount of data. However, for the initial steps towards becoming a data scientist, Excel is a good application with features that make it easier for researchers to conduct data science activities. Data that can be managed by Excel is not more than 1 million rows, as Excel only has a maximum of 1,048,576 rows and 16,384 columns. Nevertheless, the features in Excel are already powerful, such as error detection, removing duplicate data, correcting error values, detecting outliers, handling missing data, and validating data. This study discusses the functions of these features in an effort to promote data science for beginner data scientists..*

**Keywords:** *data science; data preparation; excel*

## PENDAHULUAN

Data cleaning adalah istilah yang dapat ditemui dalam data science (Setiawan, 2021). Data cleaning sendiri sendiri adalah pemahaman tentang peningkatan kualitas data. Ilmuan data sebelum melakukan pemilihan model, akan menghabiskan 50% waktunya untuk melakukan persiapan data, karena kualitas data sangat memengaruhi hasil analisis. Data cleaning merupakan prosedur yang memastikan kualitas sebuah data.

Pembersihan data merupan tahapan-tahapan yang bertujuan memastikan keakuratan, konsistensi, dan kegunaan data dalam kegiatan pengumpulan data (Pandita et al., 2018)(Ruel et al., 2018). Rahasianya adalah mendeteksi kesalahan atau korupsi data dan memperbaiki atau menghapus data sesuai kebutuhan. Menggabungkan beberapa sumber data pada saat yang sama dapat menghasilkan data duplikat atau salah label. Dalam situasi ini, sanitasi data juga diperlukan untuk menghindari masalah yang lebih kompleks.

Ada beberapa alasan mengapa pembersihan data (data cleaning) merupakan proses yang penting dilakukan. Salah satunya adalah untuk menghilangkan kesalahan dan inkonsistensi yang muncul ketika beberapa sumber data dikumpulkan menjadi satu set data. Proses ini juga dapat meningkatkan efisiensi kerja karena memudahkan para pengembang dan tim pengolah data dalam menemukan informasi yang diharapkan berdasarkan data. Dengan tingkat kesalahan yang lebih rendah, pelanggan akan merasa lebih puas dan beban kerja tim dapat berkurang. Selain itu, proses pembersihan data juga membantu para pengembang dalam memetakan beberapa fungsi data yang berbeda. Dengan demikian, proses ini dapat membantu para pengembang lebih memahami penggunaan data dan mengetahui asal-usul data tersebut.

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*

## METODE PENELITIAN

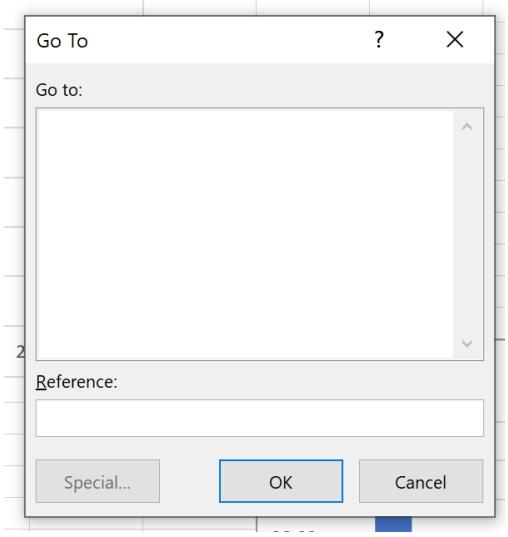
Metode data cleaning yang dilakukan dengan excel pada kegiatan ini meliputi (Hossain, 2021) :

### 1 Mendeteksi Error

Langkah awal yang perlu dilakukan adalah melihat pesan kesalahan atau korupsi data (error) (Huang & He, 2018). Terkadang, kesalahan tidak terlihat saat mengedit dokumen, tetapi muncul ketika mencetak dokumen. Ilmuan data sebenarnya tidak perlu melihat setiap sel pada worksheet untuk menemukan sel yang bermasalah. Microsoft Excel menyediakan fasilitas "Go To" yang dapat dipakai untuk menemukan kesalahan tersebut.

Untuk menemukan sel error yang masih terdapat kesalahan dapat dilakukan dengan Langkah sebagai berikut :

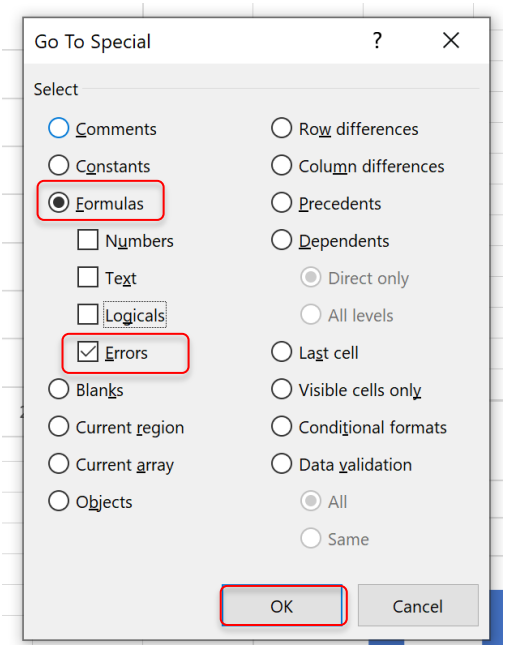
a. Block sel data yang akan dicari errornya, lalu Tekan F5 atau CTRL+G sebagai shortcut




**Gambar 1**. Visualisasi Fitur Go To

b. Pada kotak dialog Go To yang muncul, klik tombol Special
c. Kotak Go To Special akan muncul, selanjutnya pilih opsi Formulas dan centang pada menu “Errors”

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*



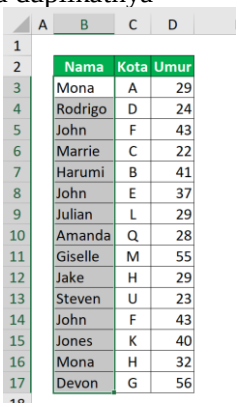

**Gambar 2**. Visualisasi Tampilan Go To Spesial

d. Akhiri dengan menekan OK
e. Sel yang teridikasi error akan diblok dan dapat ditandai dengan memberikan warna, missal warna kuning (Yellow).

## 2 Hapus Duplikat Data Atau Data Yang Tidak Perlu

Conditional Formatting pada Microsoft Excel adalah suatu fitur yang dapat digunakan untuk memberikan format khusus pada sel-sel yang memenuhi aturan tertentu pada suatu range atau table (Wang & He, 2019). Dengan menggunakan fitur ini, pengguna dapat lebih mudah melakukan analisis data dalam jumlah besar. Cara paling mudah untuk mengetahuai data double di excel atau data ganda dan duplikat pada excel adalah dengan menggunakan fitur conditional formatting.

a. Block sel yang ingin di cari data duplikatnya

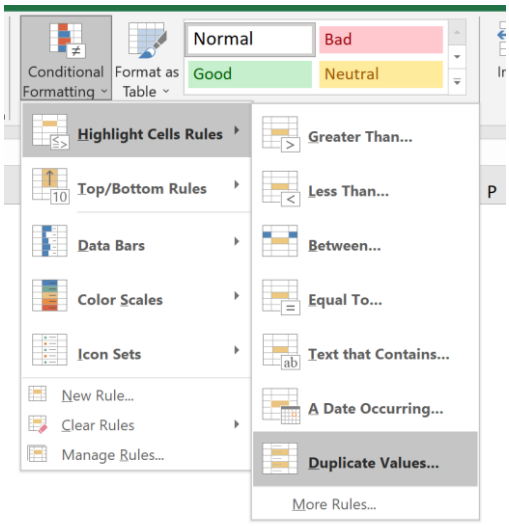


| | A | B | C | D |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | Nama | Kota | Umur |
| 3 | | Mona | A | 29 |
| 4 | | Rodrigo | D | 24 |
| 5 | | John | F | 43 |
| 6 | | Marrie | C | 22 |
| 7 | | Harumi | B | 41 |
| 8 | | John | E | 37 |
| 9 | | Julian | L | 29 |
| 10 | | Amanda | Q | 28 |
| 11 | | Giselle | M | 55 |
| 12 | | Jake | H | 29 |
| 13 | | Steven | U | 23 |
| 14 | | John | F | 43 |
| 15 | | Jones | K | 40 |
| 16 | | Mona | H | 32 |
| 17 | | Devon | G | 56 |

**Gambar 3.** Seleksi Rentang Data

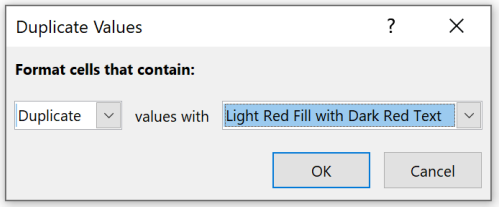
b. Selanjutnya klik menu home → Conditional Formatting → Highlights Cells Rules→ Duplicates Values

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*




**Gambar 4.** Visualisasi Menu Duplicate Value

c. Tentukan format cel jika terdapat data yang duplikat akan di rubah seperti apa. Pada gambar dibawah sel akan dirubah menjadi mereah pada data yang memiliki duplikat.




**Gambar 5.** Setting Property Nilai Yang Duplicate

d. Akhiri dengan meng-klik tombol ok.

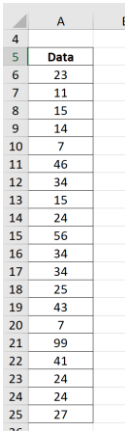
### 3 Perbaiki Kesalahan Struktur

Excel pada dasarnya mengelompokan data menjadi numerik dan text. Numerik berada di sebelah kanan cell sedangkan teks ada disebelah kiri cell. Data yang masuk dalam kategori numerik maka dapat dilakukan operasi bilangan (Liu et al., 2018). Sedangkan teks tidak dapat dilakukan operasi bilangan. Contoh pada data nomor telepon, walaupun nilainya berupa angka tapi tidak dapat dilakukan operasi bilangan seperti dikali, dibagi, ditambah atau di kurangi. Olehk karena itu data nomor telepon digolongkan sebagai data text. Untuk memperbaiki kesalah struktur data dapat menggunakan 2 fungsi yaitu replace dan subtitue.

### 4 Filter Outlier Yang Tidak Diinginkan

Pada proses pengolahan data, terkadang muncul data yang tampaknya tidak sesuai atau jauh berbeda dengan data lainnya, yang disebut sebagai outlier atau pencilan (Wu et al., 2020). Jika ditemukan outlier, mungkin dapat dihapus, tetapi harus ada alasan yang jelas. Penghapusan outlier dapat membantu meningkatkan performa data yang sedang dikerjakan. Namun, penting diingat bahwa keberadaan outlier tidak selalu menunjukkan kesalahan pada teori yang sedang dikerjakan (Grech, 2018). Sebaliknya, keberadaan outlier dapat digunakan sebagai indikator untuk memeriksa validitas data. Berikut adalah contoh data yang perlu diperiksa keberadaan outlier-nya.

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*

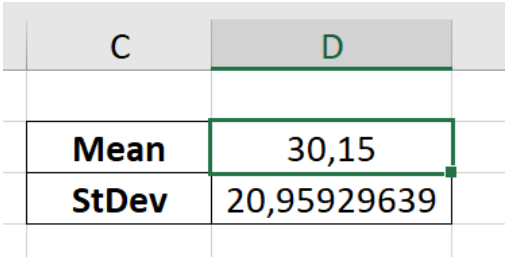


| | A |
|---|---|
| 4 | |
| 5 | Data |
| 6 | 23 |
| 7 | 11 |
| 8 | 15 |
| 9 | 14 |
| 10 | 7 |
| 11 | 46 |
| 12 | 34 |
| 13 | 15 |
| 14 | 24 |
| 15 | 56 |
| 16 | 34 |
| 17 | 34 |
| 18 | 25 |
| 19 | 43 |
| 20 | 7 |
| 21 | 99 |
| 22 | 41 |
| 23 | 24 |
| 24 | 24 |
| 25 | 27 |

**Gambar 6.** Data yang akan di lakukan pengecekan outlier

Berikut adalah langkah-langkah untuk melakukan deteksi outlier pada data:

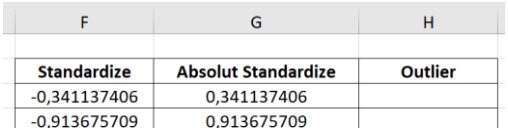
a. Hitung rata-rata atau Mean dari kumpulan data di kolom D5 menggunakan rumus: =average(A6:A25).
b. Hitung nilai standar deviasi dari data di kolom D6 menggunakan rumus: =Stdev.s(A6:A25).



| | C | D |
|---|---|---|
| | **Mean** | 30,15 |
| | **StDev** | 20,95929639 |

**Gambar 7.** Hasil pencarain nilai standar deviasi

c. Beri label pada kolom F5, G5, dan H5 dengan urutan: Standarisasi, Standarisasi Mutlak, dan Outlier.



| F | G | H |
|---|---|---|
| **Standardize** | **Absolut Standardize** | **Outlier** |
| -0,341137406 | 0,341137406 | |
| -0.913675709 | 0.913675709 | |

**Gambar 8.** Pelabelan data

d. Hitung nilai standarisasi dari setiap data pada kolom A6:A25 di kolom F6:F25 menggunakan rumus: =STANDARDISASI(A6, D$5, D$6). Nilai ini akan menghasilkan data terstandarisasi berdasarkan nilai Mean dan Standar Deviasi pada langkah a dan b.
e. Salin rumus pada sel F6 dan tempelkan pada sel F7 hingga F25.
f. Hitung nilai absolut dari nilai standarisasi pada kolom F6:F25 dengan rumus: =ABS(F6).
g. Salin rumus pada sel G6 dan tempelkan pada sel G7 hingga G25.
h. Untuk menentukan apakah data merupakan outlier atau tidak, ketikkan rumus pada sel H6: =IF(G6>3,”*”,””). Jika nilai absolut dari nilai standarisasi melebihi angka 3, maka data tersebut dianggap sebagai outlier dan akan ditandai dengan tanda * di kolom H.
i. Salin rumus pada sel H6 dan tempelkan pada sel H7 hingga H25.
j. Periksa hasil pada kolom H7:H25. Jika terdapat tanda *, maka data tersebut dianggap sebagai outlier.

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*

| F | G | H |
|---|---|---|
| **Standardize** | **Absolut Standardize** | **Outlier** |
| -0,341137406 | 0,341137406 | |
| -0,913675709 | 0,913675709 | |
| -0,722829608 | 0,722829608 | |
| -0,770541134 | 0,770541134 | |
| -1,104521811 | 1,104521811 | |
| 0,756227676 | 0,756227676 | |
| 0,183689372 | 0,183689372 | |
| -0,722829608 | 0,722829608 | |
| -0,293425881 | 0,293425881 | |
| 1,233342929 | 1,233342929 | |
| 0,183689372 | 0,183689372 | |
| 0,183689372 | 0,183689372 | |
| -0,245714355 | 0,245714355 | |
| 0,6130931 | 0,6130931 | |
| -1,104521811 | 1,104521811 | |
| 3,284938517 | 3,284938517 | * |
| 0,517670049 | 0,517670049 | |
| -0,293425881 | 0,293425881 | |
| -0,293425881 | 0,293425881 | |
| -0,150291305 | 0,150291305 | |

**Gambar 9.** Hasil pendeteksian nilai yang outlier

## 5 Penanganan Data Yang Hilang

Missing Value merujuk pada data yang hilang dalam proses pengolahan data. Dalam data science, missing value penting dalam proses persiapan data (data wrangling) sebelum melakukan analisis dan prediksi data (Kaminskyi et al., 2018) (Biessmann et al., 2018). Data wrangling adalah proses membersihkan data dari format yang tidak seragam, missing value, dan tambahan sufiks atau prefiks. Seorang data scientist dapat menghabiskan waktu 60% untuk proses ini karena 75% data perusahaan dianggap kotor. Salah satu teknik yang dapat digunakan untuk memperbaiki missing value pada Excel adalah dengan menggunakan teknik interpolasi linier.

K11

| | A | B | C | D |
|---|---|---|---|---|
| 1 | interpolasi | | nilai a | |
| 2 | | | 14 | |
| 3 | | | 15 | |
| 4 | | | 16 | |
| 5 | | | 17 | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | 21 | |
| 10 | | | 22 | |
| 11 | | | 23 | |
| 12 | | | 24 | |

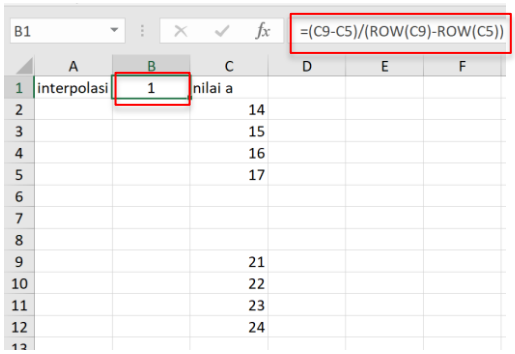
**Gambar 10.** Posisi nilai yang hilang atau missing

Cell B6, B7 dan B8 nilainya hilang atau tidak terisi. Kemungkinan penyebabnya adalah salah input atau system tidak divalidasi sehingga nilai pada kolm tersebut dapat diisi kosong. Untuk mengatasi keadaan tersebut dapat menggunakan rumus

(1) =(angkaTerbesar–angkaTerkecil)/ (ROW(angkaTerbesar)– ROW(angkaTerkecil)

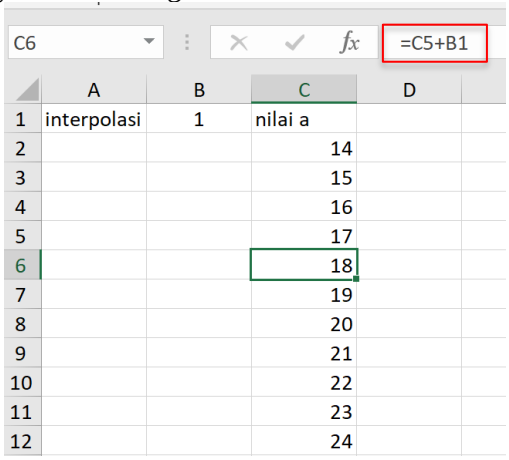
Bila dimasukan dalam excel rumus tersebut adalah sebagai berikut

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*




**Gambar 11.** Posisi angkan interpolasi setelah dicari menggunakan rumus

Selanjutnya masukan nilai untuk cell c6 dengan rumus c5 + B1 (nilai interpolasi yang ditemukan), maka hasilnya adalah sebagai berikut




**Gambar 12.** Proses menginputkan angka yang missing dengan menjumlahkan anga terkecil dengan nilai interpolasi yang didapatkan

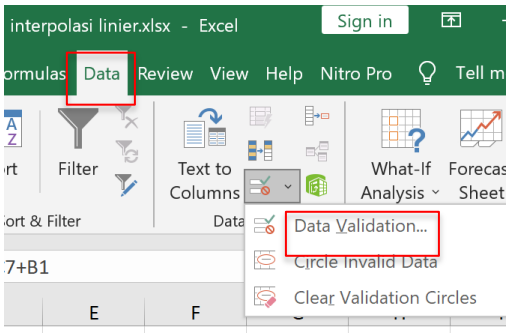
## 6 Validasi Data

Langkah terakhir dalam pembersihan data adalah validasi. Di Microsoft Excel, validasi data memungkinkan untuk membatasi nilai atau teks yang diinput pada sebuah sel atau range dengan kriteria tertentu. Misalnya, sebuah sel hanya boleh diisi dengan angka 1-10, hanya boleh menggunakan daftar teks tertentu, atau hanya dapat diisi dengan format tanggal. Untuk kebutuhan seperti ini, Microsoft Excel menyediakan fitur yang disebut "Validasi Data" atau "Validasi Data Excel" (Sofalvi & Schueler, 2021).

Berikut adalah cara untuk membuat validasi data di Excel untuk membatasi isi sebuah sel atau range:

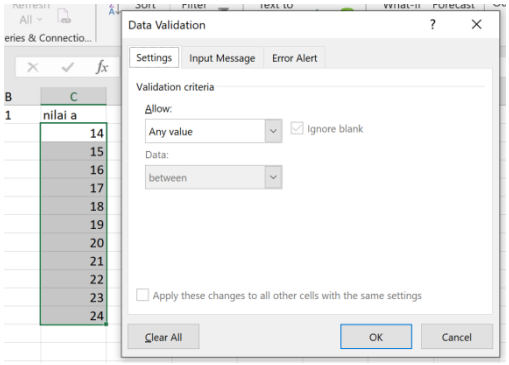
a. Pilih sel/range yang akan diatur validasi datanya.
b. Pilih menu Validasi Data pada tab Data - Group Data Tools.

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*



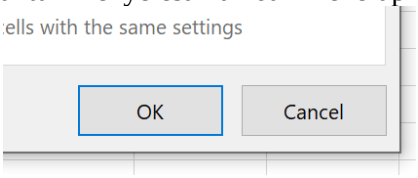

**Gambar 13.** Menu data validation

c. Langkah di atas juga dapat dilakukan dengan menggunakan shortcut keyboard Alt + A + V + V.
d. Setelah muncul kotak opsi Validasi Data, atur pengaturan validasi data atau pembatasan isi sel yang diinginkan.




**Gambar 14.** Visualisasi menu data validation

e. Langkah di atas juga dapat dilakukan dengan menggunakan shortcut keyboard Alt + A + V + V.
f. Setelah muncul kotak opsi Validasi Data, atur pengaturan validasi data atau pembatasan isi sel yang diinginkan.
g. Terakhir, pilih/klik OK untuk menyelesaikan dan menerapkan pengaturan validasi data.




**Gambar 15.** Klik ok untuk melanjutkan

## HASIL DAN PEMBAHASAN

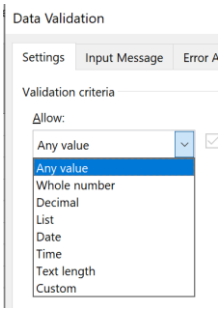
Sebagaimana yang telah dibahas sebelumnya, terdapat banyak faktor yang dapat memengaruhi hasil analisis pada data science. Salah satu faktor kunci dalam kesuksesan tersebut terletak pada tahapan awal, yaitu data preparation yang memakan waktu paling lama dan paling sulit untuk dilakukan. Data Preparation adalah proses mempersiapkan data sehingga siap diolah untuk tahap selanjutnya (Georgieva et al., 2020). Pada tahap ini, banyak masalah yang dapat muncul dan menghambat suksesnya implementasi data science, seperti kualitas data yang buruk (missing value, data duplikat, dan data yang tidak lengkap) dan dataset yang tidak

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*

seimbang (terdapat data yang terlalu dominan sehingga kegiatan data science tidak dapat melakukan prediksi dengan tepat) (Wang & He, 2019; Wu et al., 2020). Kesalahan pada tahap data preparation dapat berdampak besar pada hasil analisis data science. Selain data preparation, tahap data exploration juga merupakan hal yang tidak kalah kompleksnya. Paper ini akan difokuskan pada proses validasi data dengan menggunakan fitur-fitur yang ada pada Microsoft Excel.

## 1 Pengaturan Kriteria Pada Validasi Data

Bila dilihat lebih detail pada gambar 14 Visualisasi menu data validation, terdapat menu allow. Jika diklik maka akan memunculkan 8 menu kriteria validasi pada excel.




**Gambar 16.** Kriteria validasi data di excel

Keterangan :

a. Any Value: Sel dapat diisi dengan nilai teks atau angka tanpa batasan tertentu.
b. Whole Number: Sel hanya dapat diisi dengan angka atau bilangan bulat.
c. Decimal: Sel hanya dapat diisi dengan angka dengan batasan desimal yang dapat diatur.
d. List: Sel hanya dapat diisi dengan nilai dari daftar yang telah ditentukan sebelumnya.
e. Date: Sel hanya dapat diisi dengan nilai tanggal.
f. Time: Sel hanya dapat diisi dengan nilai waktu.
g. Text Length: Sel hanya dapat diisi dengan teks dengan batasan jumlah karakter tertentu.
h. Custom: Batasan isi sel ditentukan oleh rumus Excel atau formula yang telah ditentukan.

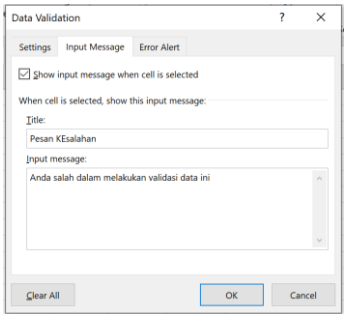
Setelah memilih salah satu dari 8 jenis data validation pada bagian Data, beberapa jenis kriteria dapat diatur lebih lanjut dengan menu-menu seperti berikut:

a. between: Hanya data yang berada di antara dua pengaturan data tertentu yang dapat dimasukkan ke dalam sel.
b. not between: Hanya data selain dari yang berada di antara dua pengaturan data tertentu yang dapat dimasukkan ke dalam sel.
c. equal to: Hanya data yang sama dengan pengaturan yang dapat dimasukkan ke dalam sel. Mirip dengan operator perbandingan "=".
d. not equal to: Hanya data yang tidak sama dengan pengaturan yang dapat dimasukkan ke dalam sel. Mirip dengan operator perbandingan "<>".
e. greater than: Hanya data yang lebih besar dari pengaturan yang dapat dimasukkan ke dalam sel. Mirip dengan operator perbandingan ">".
f. less than: Hanya data yang lebih kecil dari pengaturan yang dapat dimasukkan ke dalam sel. Mirip dengan operator perbandingan "<".
g. greater than or equal to: Hanya data yang lebih besar atau sama dengan pengaturan yang dapat dimasukkan ke dalam sel. Mirip dengan operator perbandingan ">=".
h. less than or equal to: Hanya data yang lebih kecil atau sama dengan pengaturan yang dapat dimasukkan ke dalam sel. Mirip dengan operator perbandingan "<=".

*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*

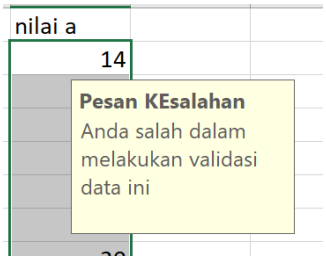
## 2 Input message pada validasi data

Tab pengaturan PESAN INPUT pada opsi "Validasi data" digunakan untuk mengatur pesan yang muncul saat sel yang diverifikasi dipilih.
Dengan mengatur Pesan Input, Excel akan menampilkan informasi tertentu saat sel yang divalidasi sedang aktif.




**Gambar 17.** Input message jika ada error

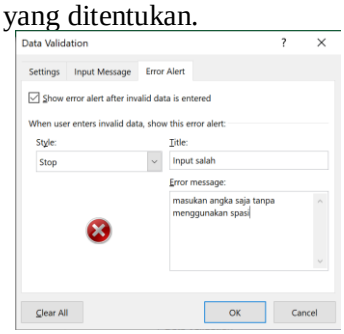
Jika opsi "Tampilkan pesan masukan saat sel dipilih" dicentang, Excel secara otomatis akan menampilkan pesan yang telah diatur di bagian Judul dan Pesan masukan saat sel yang divalidasi dipilih. Sebaliknya, biarkan judul dan pesan kosong atau hapus centang untuk tidak menampilkan pesan saat memasukkan data.
Bagian Judul adalah judul pesan yang akan ditampilkan, sedangkan Pesan Masukan adalah isi pesan yang akan muncul. Misalnya, jika validasi data diatur untuk sel A1 dan pesan masukan diatur seperti pengaturan di atas, saat sel A1 dipilih, akan muncul pesan seperti yang ditunjukkan dalam gambar di bawah ini:




**Gambar 18.** Pesan Error yang muncul saat salah dalam melakukan validasi

## 3 Error Alert Pada Data Validasi Excel

Tab pengaturan PERINGATAN KESALAHAN dari dialog "Validasi Data" digunakan untuk mengonfigurasi pesan peringatan ketika data input yang dimasukkan dalam sel yang divalidasi tidak cocok dengan pengaturan yang ditentukan.




**Gambar 19.** Pesan error

Saat pengaturan peringatan kesalahan diaktifkan, pengguna menerima informasi atau pesan peringatan jika data yang dimasukkan tidak cocok dengan kontrol data yang diterapkan.

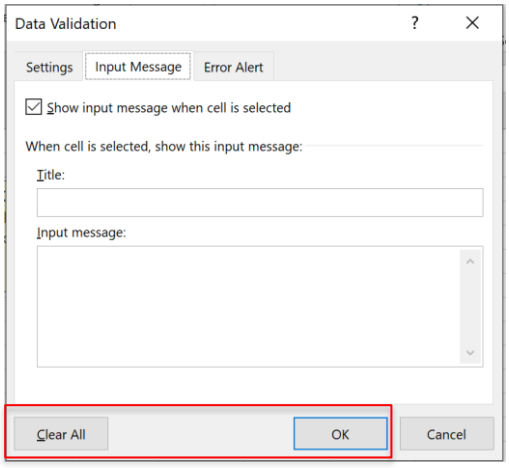
*P-ISSN: 2656-1719*
*E-ISSN: 2656-2774*

Misalnya, jika kita menyetel sel yang hanya bisa diisi dengan angka, lalu memasukkan teks ke dalam sel tersebut, peringatan kesalahan ini akan muncul secara otomatis. Jika data yang dimasukkan sudah benar sesuai dengan validasi data yang digunakan maka pesan ini tidak akan muncul.

### 4 Cara Menghapus Data Validasi Excel

Jika data validation sudah tidak dibutuhkan lagi, berikut cara untuk menghilangkannya:

a. pilih cell yang akan dihilangkan validasnya.
b. Pilih menu Data Validation yang ada pada Tab Data→Group Data Tools.
c. Klik Clear All.
d. lalu klik OK.




**Gambar 20.** Menghapus Data Validation Excel

## KESIMPULAN

Proses menjadi seorang data scientis dapat dimulai dengan mempelajari jenis data dan melakukan transofrmasi data sesuai dengan tujuan analisis. Tahapan tersebut dapat dilakukan dengan beberapa aplikasi yang akan memudahkan dan menjaga kehandalan data. Salah satu aplikasi yang dapat digunakan untuk melakukan aktifitas data scientis adalah excel. Banyak fitur-fitur excel yang dapat digunakan untuk melakukan analisis data terutama saat proses data preparation. Fungsi, rumus dan menu-menu yang ada dapat menunjang seorang data scientis dalam pekerjaannya. Untuk data yang ukurannya lebih dari 1 juta baris data sebaiknya menggunakan aplikasi lain karena excel hanya mampu menampung 1.048.576 baris dan 16.384 kolom. Tentu akan menyulitkan seorang data scientis jika harus bekerja dengan data yang lebih dari 1 juta tersebut jika menggunakan excel.

## SARAN

Kekurangan excel adalah melalui proses yang cukup banyak jika dihadapkan dengan kasus yang kompleks. Oleh karena itu perlu dibangun sebuah formula atau framework yang disepakati bersama jika ingin melakukan pengolahan data kompleks dengan excel.

**Ucapan Terimakasih**

Kami mengucapkan terimakasih kepada Universitas ‘Aisyiyah Surakarta (www.aiska-university.ac.id) melalui P3M yang telah mendanai penelitian ini.

## DAFTAR PUSTAKA

Biessmann, F., Salinas, D., Schelter, S., Schmidt, P., & Lange, D. (2018). " Deep" Learning for Missing Value Imputationin Tables with Non-numerical Data. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2017–2025.

Georgieva, P., Nikolova, E., & Orozova, D. (2020). Data Cleaning Techniques in Detecting Tendencies in Software Engineering. *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1028–1033.

Grech, V. (2018). WASP (Write a Scientific Paper) using Excel–3: Plotting data. *Early Human Development*, *117*, 110–112.

Hossain, E. (2021). MS Excel in Engineering Data. In *Excel Crash Course for Engineers* (pp. 169–242). Springer.

Huang, Z., & He, Y. (2018). Auto-detect: Data-driven error detection in tables. *Proceedings of the 2018 International Conference on Management of Data*, 1377–1392.

Kaminskyi, R., Kunanets, N., Pasichnyk, V., Rzheuskyi, A., & Khudyi, A. (2018). Recovery Gaps in Experimental Data. *COLINS*, 108–118.

Liu, R., Glover, K. P., Feasel, M. G., & Wallqvist, A. (2018). General approach to estimate error bars for quantitative structure–activity relationship predictions of molecular activity. *Journal of Chemical Information and Modeling*, *58*(8), 1561–1575.

Pandita, R., Parnin, C., Hermans, F., & Murphy-Hill, E. (2018). No half-measures: A study of manual and tool-assisted end-user programming tasks in Excel. *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 95–103.

Ruel, E., William, W., & Gillespie, B. J. (2018). Data cleaning. *The Practice of Survey Research: Theory and Applications*, 208–237.

Setiawan, I. (2021). Perbedaan Data Engineer, Data Scientist Dan Data Analyst. *Widya Accarya*, *12*(2), 306–309.

Sofalvi, S., & Schueler, H. E. (2021). Assessment of Bioanalytical Method Validation Data Utilizing Heteroscedastic Seven-Point Linear Calibration Curves by EZSTATSG1 Customized Microsoft Excel Template. *Journal of Analytical Toxicology*, *45*(8), 772–779.

Wang, P., & He, Y. (2019). Uni-detect: A unified approach to automated error detection in tables. *Proceedings of the 2019 International Conference on Management of Data*, 811–828.

Wu, Z., Wu, Z., & Rilett, L. R. (2020). Innovative nonparametric method for data outlier filtering. *Transportation Research Record*, *2674*(10), 167–176.