**TUGAS DATA MINING**

**PERTEMUAN MINGGU KE – 2**

1. Carilah data yang dapat Anda gunakan untuk Proses Data Mining kemudian gunakan CRISP DM untuk menyelesaikan masalah tersebut !
2. Jelaskan setiap langkahnya seperti Studi Kasus Data Kelulusan Mahasiswa di atas menggunakan proses CRISP DM dimulai dari Business Understanding Data Understanding Data Preparation Modeling Evaluation dan Deployment !

**Jawaban :**

1. Data yang digunakan untuk proses data mining pada kasus ini adalah data pemilihan umum KPU yang berekstensi .xls.
2. Proses CRISP DM :

* **Pemahaman Proses Bisnis / Bussiness Understanding**

1. Problem

* Mbak PM adalah seorang anggota / kader dari partai P dari daerah tertentu.
* Sebagai seorang kader, tentunya terjun di dalam dunia politik baik legislatif maupun eksekutif merupakan keinginnya.
* Pada pemilihan umum (pemilu) legislatif 2024 mendatang, Mbak PM ingin mencalonkan dirinya sebagai calon legislatif (caleg).
* Namun, Mbak PM memiliki masalah karena dirinya merasa tidak yakin bisa memenangi atau memprediksi tingkat keterpilihan dirinya di dalam pemilu tersebut.
* Oleh karena itu, Mbak PM ingin memahami dan membuat pola dari profil caleg sebelumnya yang bisa terpilih dan tidak menjadi anggota legislatif di daerahnya.
* Dengan menggunakan pola tersebut, Mbak PM dapat melakukan prediksi tingkat keterpilihannya sehingga dia tidak ragu lagi membuat keputusan untuk tetap mencalonkan sebagai caleg dan pasti terpilih jika tingkat keterpilihannya tinggi atau mengurungkan niatnya agar terhindar dari kerugian materi imbas dari kampanye pemilu jika tetap nekat mencalonkan diri sedangkan peluang untuk memenangi pemilu di daerahnya tipis.

1. Objektif

Menemukan pola dari calon legislatif yang terpilih dan tidak.

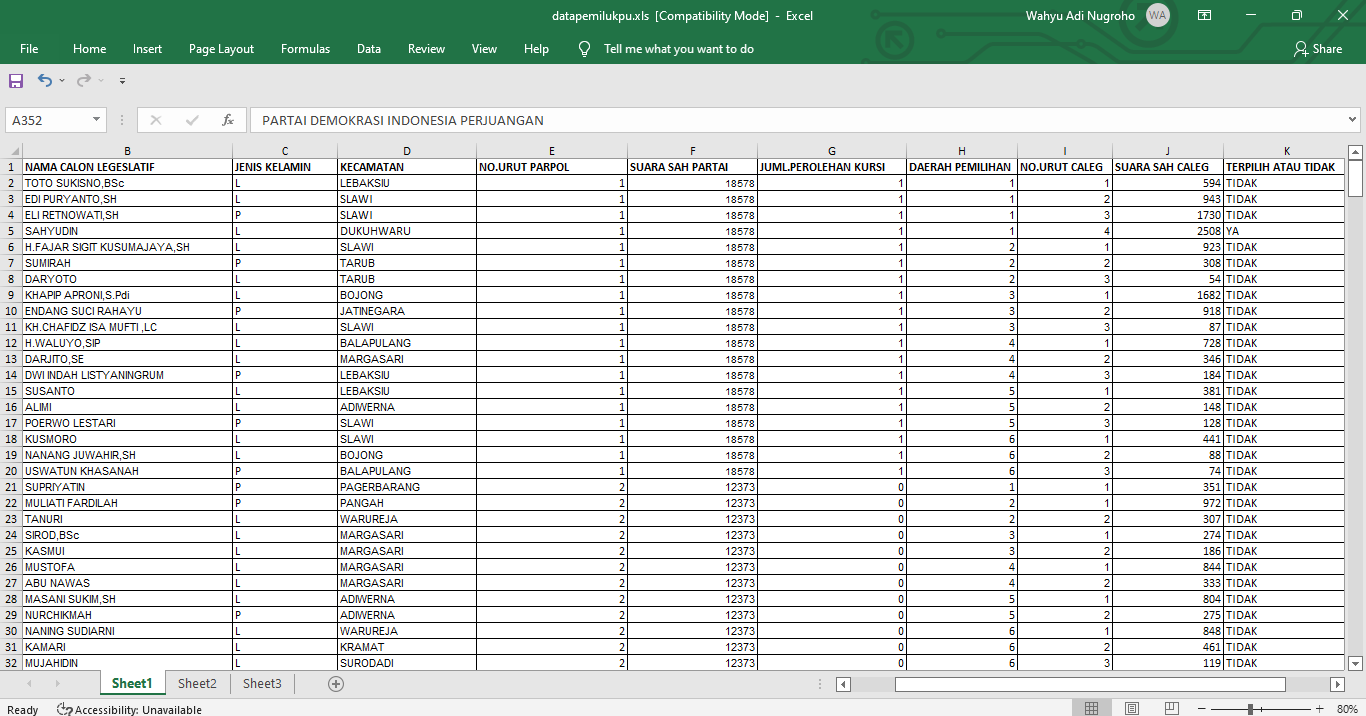
* **Pemahaman Data / Data Understanding**

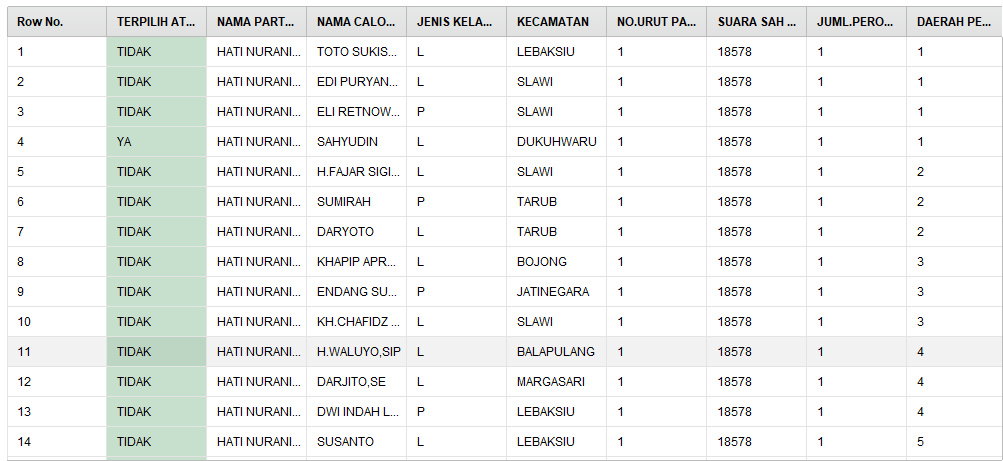
Untuk meyelesaikan masalah tersebut, Mbak PM mengambil dari dari website Situng KPU. Data – data dikumpulkan dari data pemilihan umum kpu dengan atribut seperti berikut.

1. NAMA PARPOL.
2. NAMA CALON LEGISLATIF.
3. JENIS KELAMIN : L untuk laki – laki dan P untuk Perempuan.
4. KECAMATAN.
5. NO. URUT PARPOL.
6. SUARA SAH PARTAI.
7. JUML. PEROLEHAN KURSI.
8. DAERAH PEMILIHAN.
9. NO. URUT CALEG.
10. SUARA SAH CALEG.
11. STATUS KETERPILIHAN : Terpilih atau Tidak.

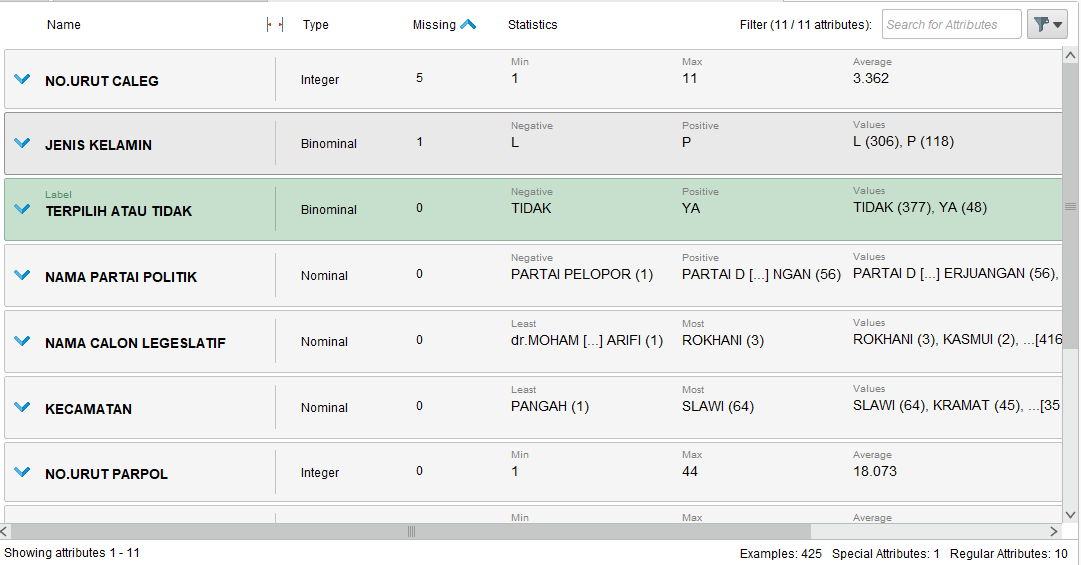
* **Persiapan Data / Data Preparation**

1. Dataset : datapemilukpu.xls

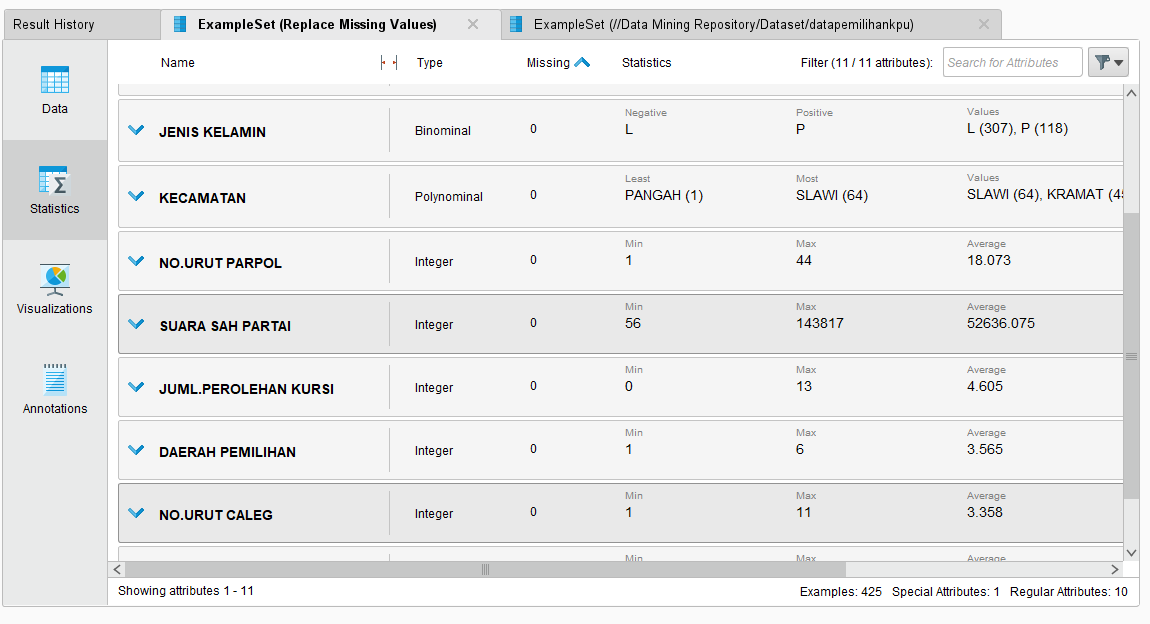




1. Terdapat 425 data caleg dengan 11 atribut (1 special attributes/label, 10 regular attributes) dengan missing value sebanyak 6 buah dan tidak terdapat data noise.

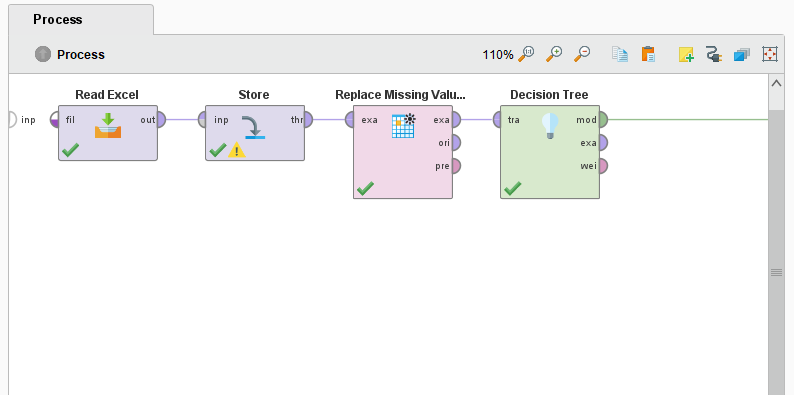


1. Missing Value dipecahkan dengan menambahkan data dengan nilai rata rata. • Hasilnya adalah data bersih tanpa missing value.

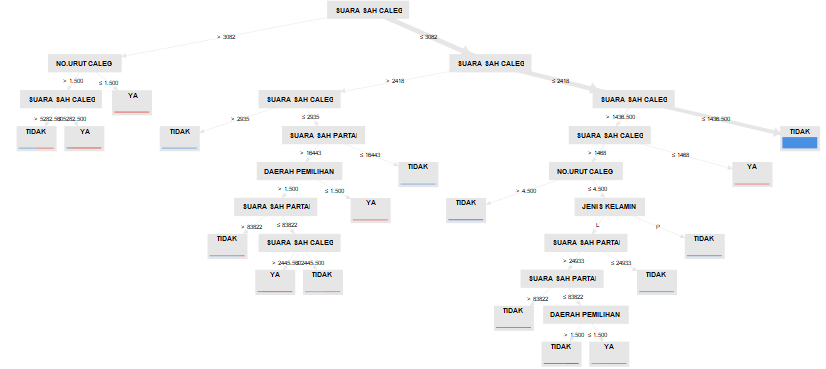


* **Pemodelan / Modeling**

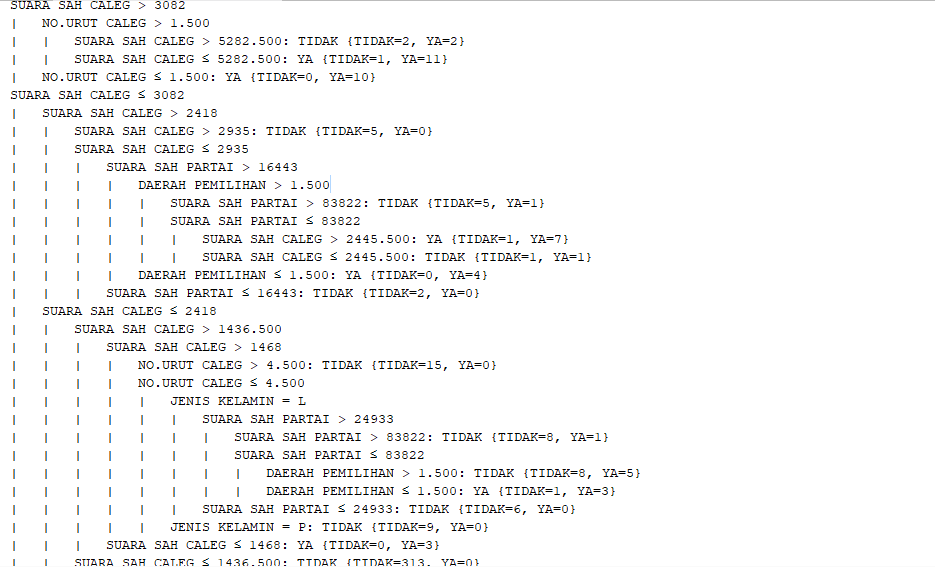
Memodelkan dengan Decision Tree.



Hasil pola dari data berupa berupa decision tree pohon keputusan sebagai berikut.



* **Evaluasi / Evaluation**

****

Dari deskripsi di atas, dapat diketahui bahwa :

1. Atribut atau faktor yang paling berpengaruh adalah : **SUARA SAH CALEG, NO URUT CALEG.**
2. Atribut atau faktor yang tidak berpengaruh adalah : **DAERAH PEMILIHAN, SUARA SAH PARTAI, JENIS KELAMIN.**

* **Penerapan / Deployment**

1. Mbak PM akan melakukan strategi kampanye yang bersih dengan menawarkan program kerja yang pro rakyat sehingga diharapkan dia dapat mendapatkan suara yang mayoritas karena faktor yang menentukan keterpilihan dirinya sebagai anggota legislatif ialah surat suara caleg yang sah.
2. Mbak PM juga akan mengusahakan untuk mendapatkan no urut caleg yang tingkat keterpilihannya tinggi kendati pun pada kenyataannya akan susah untuk mendapatkan no urut sesuai dengan keinginannya karena biasanya pemilihan no urut berdasarkan undian sehingga tergantung keberuntungan.
3. Jika kemungkinan pemilihnya kurang dari 1436.5 (dibulatkan menjadi 1437) surat suara sah, maka sebaiknya Mbak PM tidak usah mencalonkan diri sebagai caleg karena pasti **TIDAK TERPILIH**.
4. Untuk daerah pemilihan, mungkin Mbak PM perlu untuk membuat pemetaan daerah mana saja yang sekiranya kuat bagi dirinya terpilih.