

BAB IV

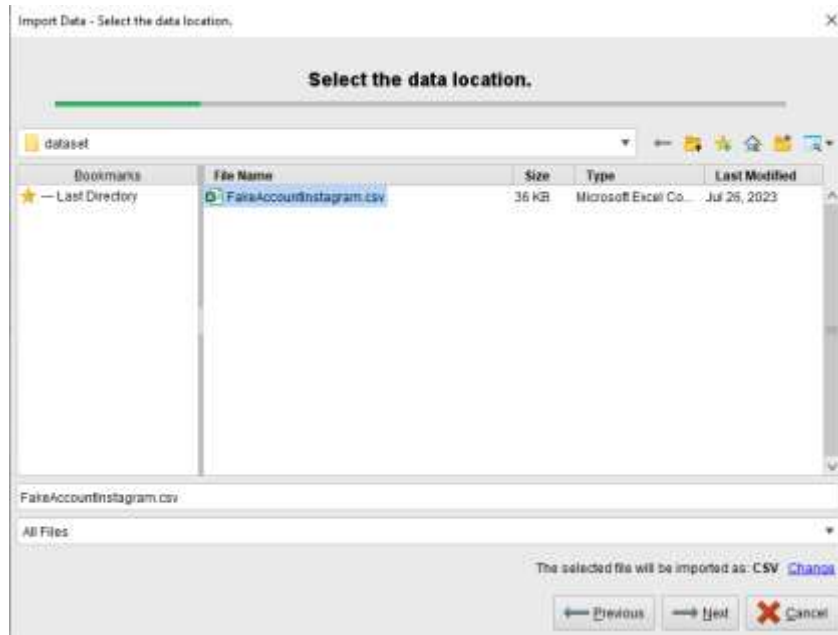
HASIL DAN PEMBAHASAN

4.1. Hasil dan Pembahasan

Pada penelitian ini digunakan ini menggunakan prediksi dengan algoritma Decision Tree C4.5 dan Adaboost. Topik yang di ambil dalam penelitian ini adalah prediksi fake account pada media social Instagram dengan data sampling sebesar 1.152 data yang terkumpul. Data tersebut diolah dengan melakukan preprocessing sehingga siap untuk dilakukan pengolahan dengan data mining. Pada modeling data menggunakan salah satu algoritma dari metode machine learning yaitu decision tree dimana penentuan prediksi akun dapat dilihat dengan menggunakan pohon keputusan lalu dilakukan validasi hasil dengan dilakukan boosting menggunakan adaboost untuk meningkatkan performa dari data tersebut. Data sample ini juga dihitung nilai probabilitasnya, kemudian akan dicari tingkat akurasi tertinggi dari data tersebut dan akan di implementasikan pada tools RapidMiner.

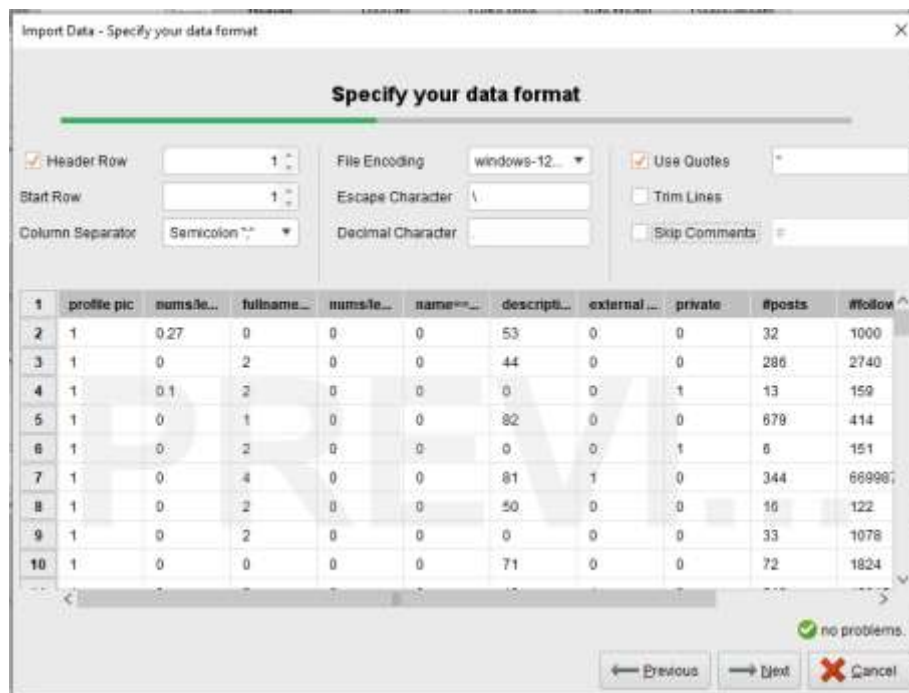
4.1.1. Hasil Pengolahan Data

Pada hasil pengolahan data dalam dataset merupakan hasil dari pemrosesan saat dataset siap digunakan untuk dilakukan tahapan preprocessing. Tool yang digunakan dalam tahapan ini adalah Rapidminer kemudian dataset yang digunakan merupakan data excel dengan format csv. pada tahapan pengimporan dataset dilakukan dengan memilih file import pada tool rapidminer kemudian akan ada menu tampilan pada import data. Data yang diimport adalah data awal yang berasal dari data publik sehingga setelah melakukan pengimporan akan dilanjutkan pada proses preprocessing. Berikut adalah tahapan tahapan dalam import dataset.



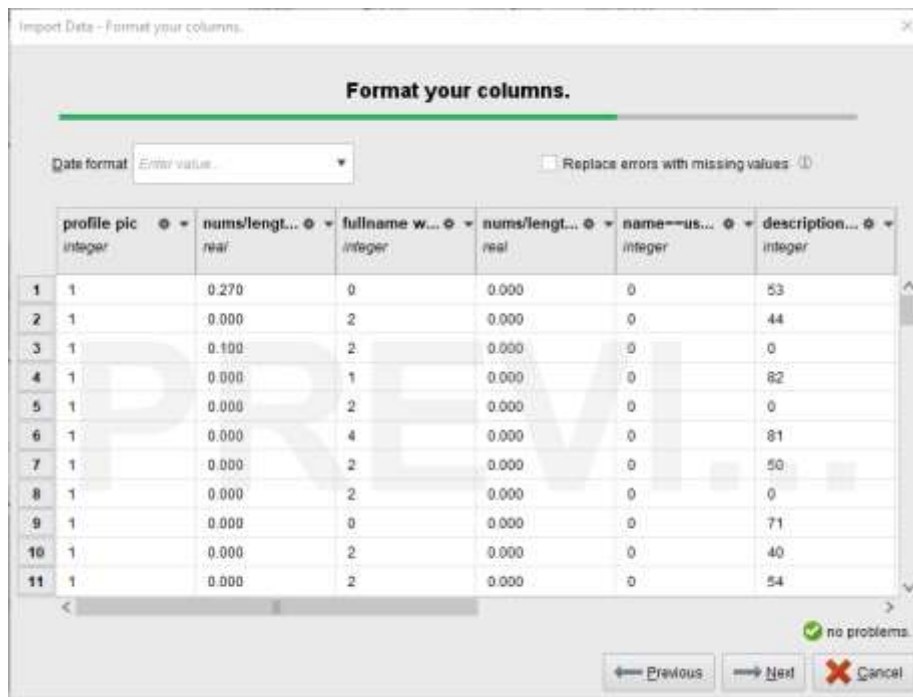
Gambar 4. 1 Proses Import Dataset

Setelah tahapan pengimporan selanjutnya adalah tampilan halaman spesifik dalam dataset. Dalam tahapan ini dilakukan untuk memastikan apakah format yang digunakan sudah benar sesuai dengan atribut yang kita gunakan.



Gambar 4. 2 Format Spesifik Data

Setelah format data maka akan muncul tampilan untuk dilakukan format pada kolom data pada dataset. Pada proses ini akan dilakukan perubahan tipe atribut sesuai dengan karakteristik atribut yang dimiliki. Dalam proses ini adalah tahapan perubahan untuk menentukan tipe data apa yang akan digunakan pada setiap atribut. Pada tahapan ini menggunakan 3 tipe atribut yang akan digunakan yaitu tipe data integer, real dan binomial dimana pada dataset yang digunakan umumnya bersifat nilai value dan juga terdapat nilai data pada atribut yang bersifat koma menggunakan data real dan data label yang digunakan menggunakan tipe binomial. Pada tahapan ini akan sangat mempengaruhi hasil dari performa yang akan kita jalankan. Berikut adalah tampilan dari format pada tipe data pada setiap atribut.



Gambar 4. 3. Pelabelan Tipe Data Atribut

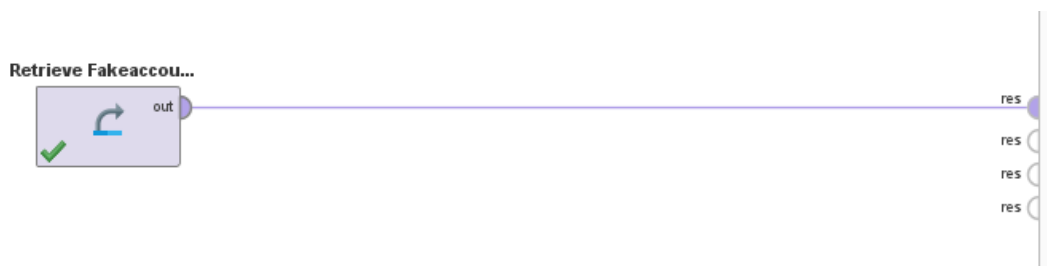
Pada gambar diatas dapat kita lihat adalah beberapa perubahan atau pelabelan pada data tiap-tiap atribut dalam dataset. Dalam tahapan ini hanya melakukan pelabelan pada tiap-tiap atribut saja. Berikut adalah tipe atribut setelah dilakukan pelabelan pada tipe data dalam dataset yang digunakan dalam bentuk table sesuai

pada pelabelan data dalam tool rapidminer.

Tabel 4. 1 Tipe Data Atribut

| No | Nama Atribut | Tipe Data |
|----|----------------------|-----------|
| 1 | profile pic | binominal |
| 2 | nums/length username | real |
| 3 | fullname words | integer |
| 4 | nums/length fullname | real |
| 5 | name username | binominal |
| 6 | description length | integer |
| 7 | external URL | binominal |
| 8 | private | binominal |
| 9 | posts | integer |
| 10 | followers | integer |
| 11 | follows | integer |
| 12 | fake | binominal |

Pada tampilan dataset yang sudah di import dapat dilihat dengan menarik langsung hasil dari file yang di import ke dalam lembar kerja di dalam tool rapidminer seperti pada hasil gambar berikut.



Gambar 4. 4 Tampilan Dataset pada Rapidminer

Pada tampilan diatas merupakan logic dari dataset dan jika di running akan menampilkan statistic dalam tool. Dimana dalam statistik data terseut dapat dilihat missing data dan juga data fake akun positive dan negative pada dataset tersebut. Berikut adalah hasil tampilan statistic pada dataset yang dijalankan.



Gambar 4. 5 Statistik Dataset

Pada statistic dataset diatas dapat dilihat tidak terdapat missing data yang di dapatkan pada dataset kemudian terdapat nilai minimal dan maksimal pada setiap value dalam dataset tersebut.

4.1.2. Deskripsi dan Visualisasi Atribut

Dalam atribut yang digunakan sebagai parameter penentuan prediksi fake account untuk penelitian adalah sebanyak 11 atribut. Dalam deskripsi pada atribut akan dijelaskan batas keterangan dari nilai pada tiap atribut yang digunakan yang nantinya akan di dapatkan dalam pohon keputusan dalam menentukan root awal pada prediksi fake account Instagram. Berikut adalah deskripsi dan visualisasi pada tiap-tiap atribut dalam peneltian ini.

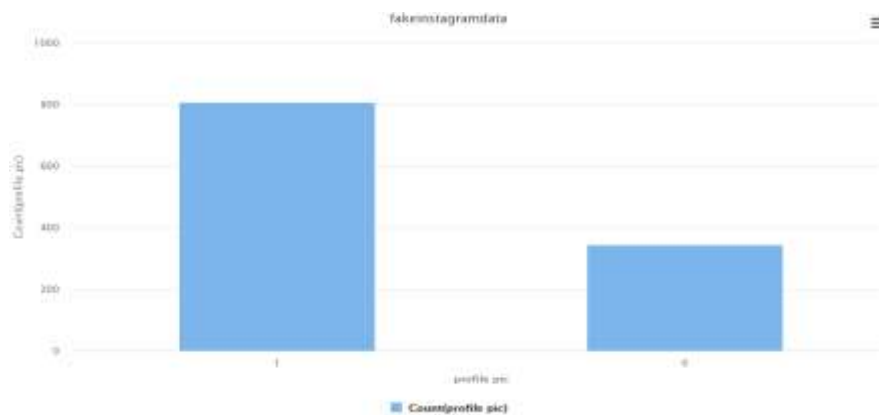
a. Profile pic

Profile pic merupakan gambar pada foto profil dari akun Instagram yang tersedia. Dalam akun tersebut ada atau tidak foto atau gambar dari profil Instagram tersebut. Dalam nilai variable yang diberikan hanya menggunakan pelabelan 1 dan 0 dengan ketentuan 1 jika terdapat foto profil dan 0 jika tidak. Berikut adalah data dari Profile pic.

Tabel 4. 2. Nominal Values pada Profile pic

| No | Nominal Value | Absolute Count | Fraction |
|----|---------------|----------------|----------|
| 1 | 1 | 808 | 0.701 |
| 2 | 0 | 344 | 0.299 |

Data table diatas merupakan jumlah dari masing masing value yang terdapat dalam Profile pic dimana hanya terdapat dua yaitu 1 dan 0. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 6 Visualisasi Profile Pic

b. Nums/Length username

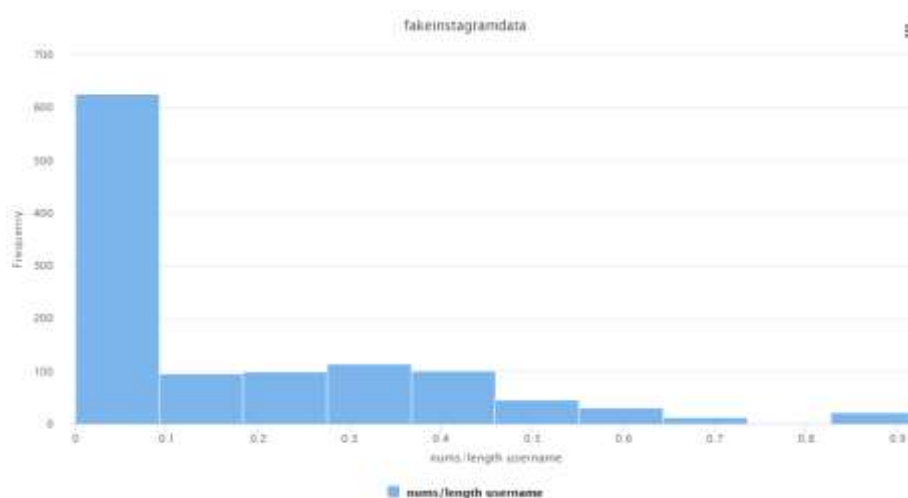
Nums/Length username adalah Panjang karakter pada nama pengguna dalam Instagram. Dalam dataset ini rentan Panjang nama pengguna adalah pada 0 sampai dengan 0.92 dalam nilai value pada atribut. Karna data value pada atribut adalah nilai real maka perhitungan jumlah data menggunakan batas yang telah di tentukan dalam pohon keputusan guna untuk menghitung nilai gain ratio pada pohon keputusan. Nilai yang akan dihitung jumlahnya adalah >0.47 dan <0.47 . berikut adalah jumlah nilainya.

Tabel 4. 3. Nominal Values pada Profile pic

| No | Rentan Nilai $><$ | Jumlah Nilai | Fake Yes | Fake No |
|----|-------------------|--------------|----------|---------|
| 1 | <0.47 | 1050 | 478 | 568 |

| | | | | |
|---|-------|-----|-----|---|
| 2 | >0.47 | 102 | 102 | 0 |
|---|-------|-----|-----|---|

Pada table diatas dapat dilihat bahwa jumlah nilai data <0.47 lebih besar dari pada nilai dengan >0.47. kemdian berikut adalah visualisasi gambar berdasarkan nilai table diatas.



Gambar 4. 7 Visualisasi Nums/Length username

c. Fullname Words

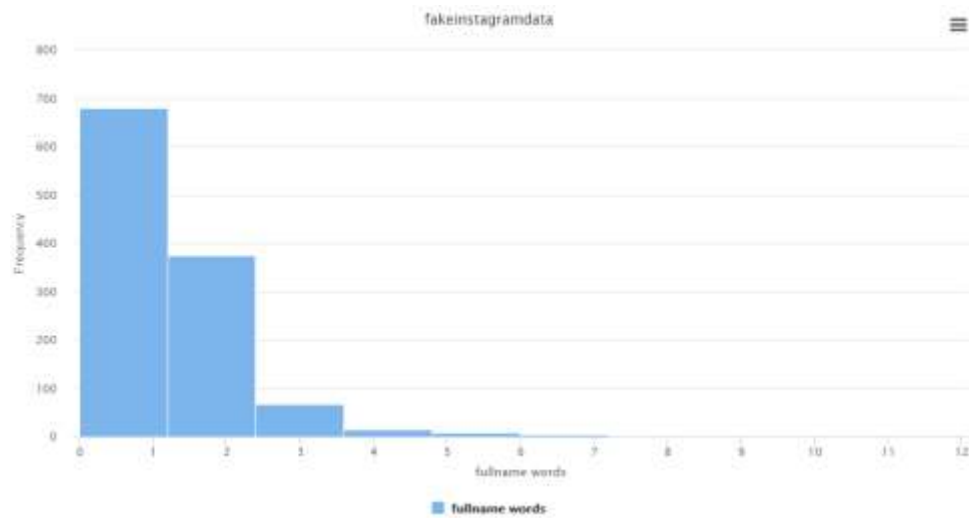
Fullname words adalah keterangan pada nama lengkap dari akun Instagram. Pada dataset penelitian ini fullname words memiliki rentan nilai dalam keterangan antara 0 sampai pada 12. Dalam pohon keputusan nilai fullname words dibagi rentan nilai antara >1 dan <1. Berikut adalah hasil nilai pada atribut tersebut.

Tabel 4. 4. Nominal Values pada Fullname words

| No | Rentan Nilai >< | Jumlah Nilai | Fake Yes | Fake No |
|----|-----------------|--------------|----------|---------|
| 1 | <1 | 680 | 472 | 208 |
| 2 | >1 | 472 | 108 | 364 |

Pada hasil table diatas dapat dilihat bahwa ternyata banyak data pada Fullname words yang kurang dari 1 daripada lebih dari 1. Berikut adalah

visualisasi gambar pada atribut fullname words.



Gambar 4. 8. Visualisasi Fullname words

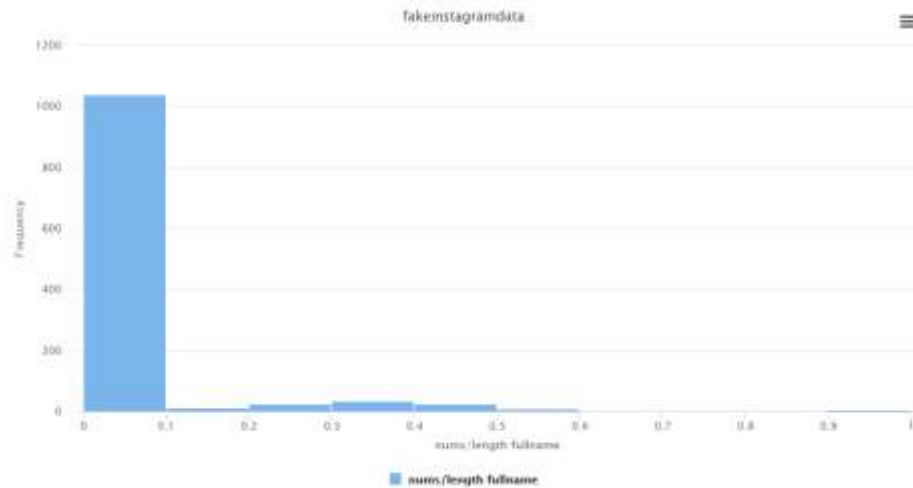
d. `nums/length fullname`

`nums/length fullname` adalah nama lengkap atau Panjang nama di akun Instagram. Dalam dataset ini Panjang nama lengkap di tentukan dengan kriteria 0 sampai 1.0. pada perhitungan nilai dibuat perbandingan dengan >0.36 dan juga <0.36 berdasarkan perbandingan pada pohon keputusan. Berikut adalah jumlah nilai pada atribut `nums/length fullname`.

Tabel 4. 5. Nominal Values pada Fullname words

| No | Rentan Nilai $<$ | Jumlah Nilai | Fake Yes | Fake No |
|----|------------------|--------------|----------|---------|
| 1 | <0.36 | 42 | 42 | 0 |
| 2 | >0.36 | 1110 | 538 | 572 |

Pada data table diatas dapat dilihat bahwa jumlah nilai pada rentan nilai >0.36 lebih besar dengan 1110 dengan perbandingan yang rata antara data fake yes dan fake no pada akun asli dan palsu. Sebaliknya nilai <0.36 justru lebih kecil dan tidak ada akun palsu pada rentan nilai ini. berikut adalah visualisasi pada atribut `nums/length`.



Gambar 4. 9. Visualisasasi Atribut nums/length fullname

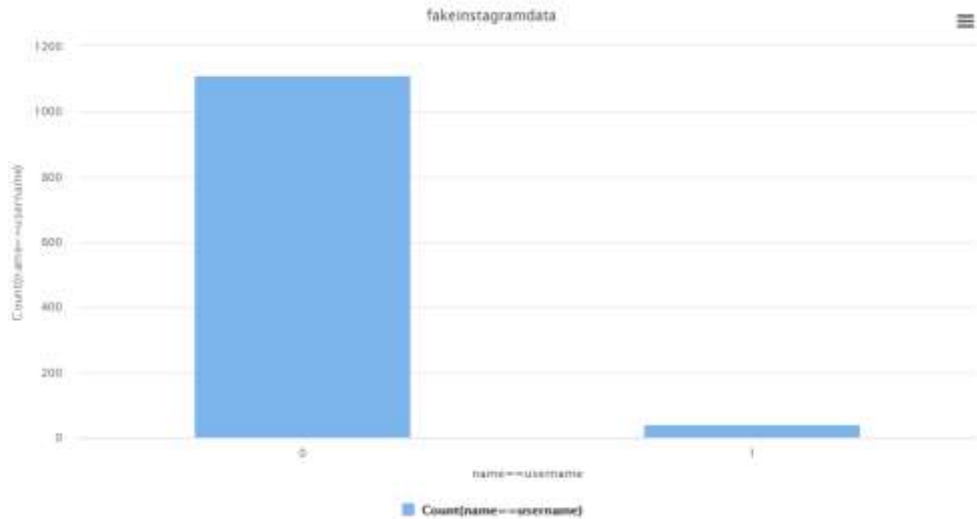
e. username

username adalah identitas yang dibuat oleh pengguna akun atau layanan online untuk membedakan dirinya dari pengguna lain. Dalam akun tersebut ada atau tidak username pengguna profil Instagram tersebut. Dalam nilai variable yang diberikan hanya menggunakan pelabelan 1 dan 0 dengan ketentuan 1 jika terdapat username dan 0 jika tidak. Berikut adalah data dari username.

Tabel 4. 6. Nominal Values pada username

| No | Nominal Value | Absolute Count | Fraction |
|----|---------------|----------------|----------|
| 1 | 1 | 40 | 0.035 |
| 2 | 0 | 1112 | 0.965 |

Data table diatas merupakan jumlah dari masing masing value yang terdapat dalam username dimana hanya terdapat dua yaitu 1 dan 0. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 10 Visualisasi Atribut username

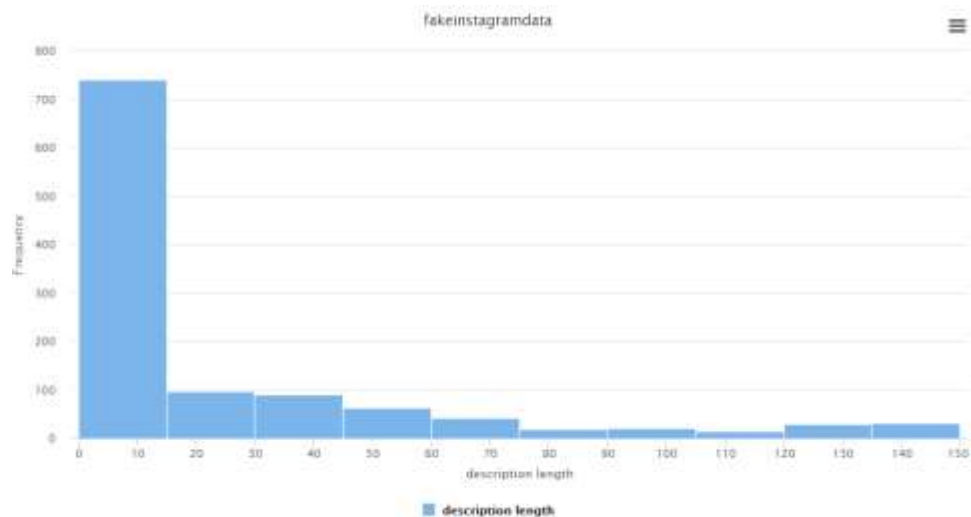
f. description length

description length adalah adalah deskripsi pada bio dalam profil Instagram dimana biasanya menyebutkan informasi tentang akun Instagram tersebut. Dalam dataset ini Panjang deskripsi akun di tentukan dengan kriteria antara 0 sampai 150 sesuai dengan batas huruf pada instagram. pada perhitungan nilai dibuat perbandingan dengan >65 dan juga <65 berdasarkan perbandingan pada pohon keputusan. Berikut adalah jumlah nilai pada atribut description length.

Tabel 4. 7. Nominal Values pada description length

| No | Rentan Nilai $><$ | Jumlah Nilai | Fake Yes | Fake No |
|----|-------------------|--------------|----------|---------|
| 1 | <65 | 1006 | 569 | 437 |
| 2 | >65 | 146 | 11 | 135 |

Pada data table diatas dapat dilihat bahwa jumlah nilai pada rentan nilai <65 lebih besar dengan 1006 dengan perbandingan yang rata antara data fake yes dan fake no pada akun asli dan palsu. Sebaliknya nilai >65 justru lebih kecil dengan jumlah nilai sebesar 146. berikut adalah visualisasi pada atribut description length.



Gambar 4. 11. Visualisasasi Atribut description length

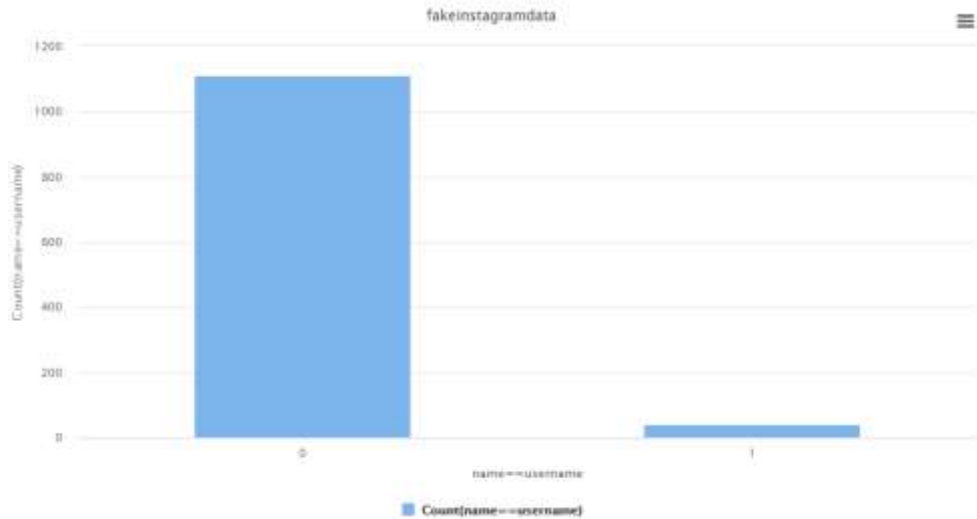
g. External URL

external URL adalah link dari halaman web atau url yang ada pada bio dalam profil Instagram. Dalam akun Instagram tak jarang para pemilik akun menautkan link url lain semisal youtube dan lain-lain. Dalam pelabelan akun tersebut nilai yang terdapat adalah ada atau tidak external URL pengguna profil Instagram tersebut. Dalam nilai variable yang diberikan hanya menggunakan pelabelan 1 dan 0 dengan ketentuan 1 jika terdapat external URL dan 0 jika tidak. Berikut adalah data dari external URL.

Tabel 4. 8. Nominal Values pada external URL

| No | Nominal Value | Absolute Count | Fraction |
|----|---------------|----------------|----------|
| 1 | 1 | 134 | 0.116 |
| 2 | 0 | 1018 | 0.884 |

Data table diatas merupakan jumlah dari masing masing value yang terdapat dalam username dimana hanya terdapat dua yaitu 1 dan 0. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 12 Visualisasi Atribut external URL

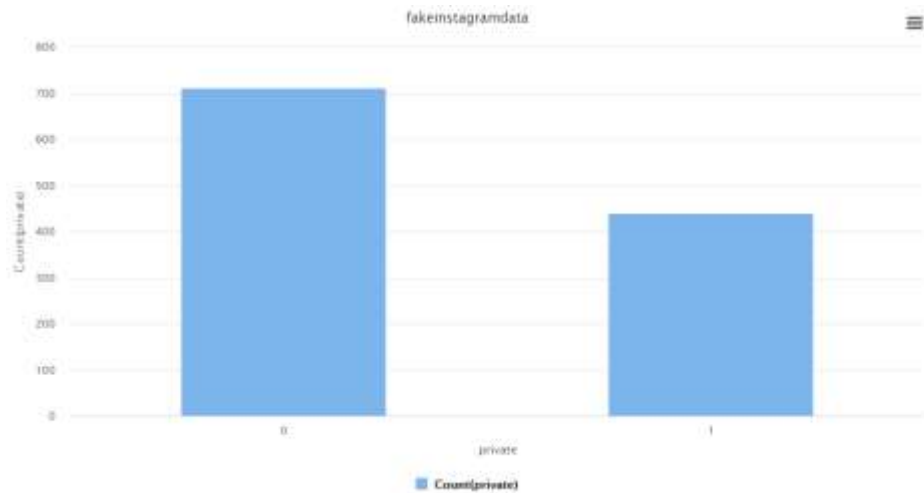
h. Private

Private adalah tipe dari akun yang di kunci atau terbuka. Akun yang terbuka artinya siapa saja yang bukan dari followers akun tersebut dapat melihat profil dari Instagram tersebut namun jika akun private adalah akun yang hanya bisa dilihat oleh orang yang mengikuti/followersnya saja. Dalam pelabelan akun tersebut nilai yang terdapat adalah jika akun tersebut di private atau tidak oleh pengguna profil Instagram tersebut. Dalam nilai variable yang diberikan hanya menggunakan pelabelan 1 dan 0 dengan ketentuan 1 jika terdapat Private dan 0 jika tidak. Berikut adalah data dari external URL.

Tabel 4. 9. Nominal Values pada external URL

| No | Nominal Value | Absolute Count | Fraction |
|----|---------------|----------------|----------|
| 1 | 1 | 440 | 0.618 |
| 2 | 0 | 712 | 0.382 |

Data table diatas merupakan jumlah dari masing masing value yang terdapat dalam atribut Private dimana hanya terdapat dua yaitu 1 dan 0. Kemudian data tersebut di visualisasikan kedalam bentuk diagram sebagai berikut.



Gambar 4. 13 Visualisasi Atribut Private

i. Post

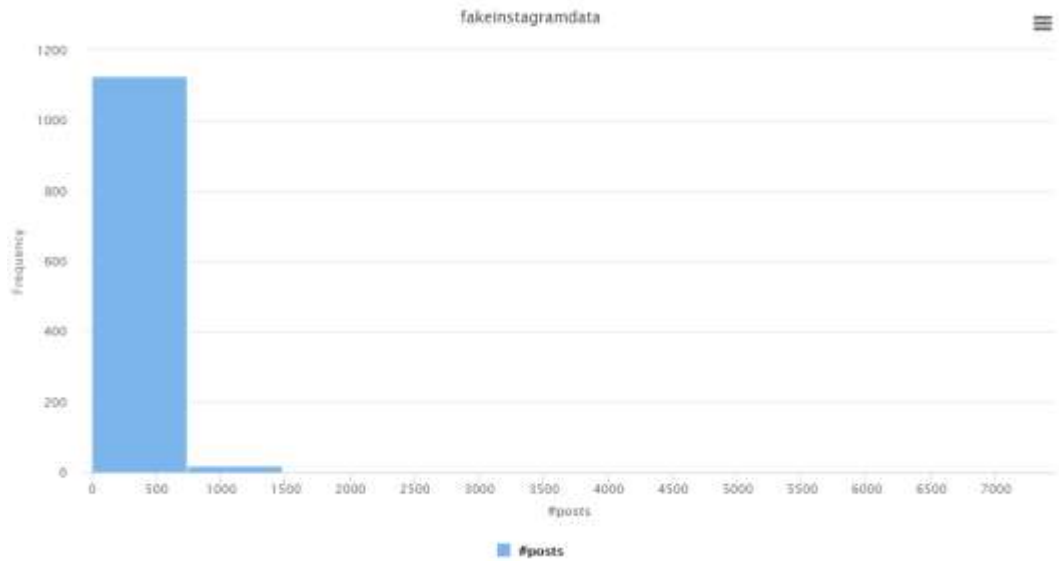
Post merupakan jumlah konten seperti foto maupun video yang di posting ke dalam profil akun Instagram tersebut. Hal ini sebagai indikator untuk memprediksi fake account. Dalam dataset ini banyaknya jumlah post dalam akun di tentukan dengan kriteria antara 0 sampai 7389 sesuai dengan jumlah tertinggi akun yang posting dalam akun Instagram. Pada perhitungan nilai dibuat perbandingan dengan >50 , <50 dan nilai 50 berdasarkan perbandingan pada pohon keputusan. Berikut adalah jumlah nilai pada atribut Post.

Tabel 4. 10. Nominal Values pada Post

| No | Rentan Nilai $<<$ | Jumlah Nilai | Fake Yes | Fake No |
|----|-------------------|--------------|----------|---------|
| 1 | <50 | 774 | 554 | 220 |
| 2 | >50 | 376 | 26 | 350 |
| 3 | 50 | 2 | 0 | 2 |

Pada data table diatas dapat dilihat bahwa jumlah nilai pada rentan nilai <50 lebih besar dengan 774 dengan perbandingan lebih besar pada data fake yes dari data fake no pada akun asli dan palsu. Sebaliknya nilai >50 justru lebih kecil dengan jumlah nilai sebesar 376. berikut adalah visualisasi pada atribut

Post.



Gambar 4. 14. Visualisasasi Atribut Post

j. Followers

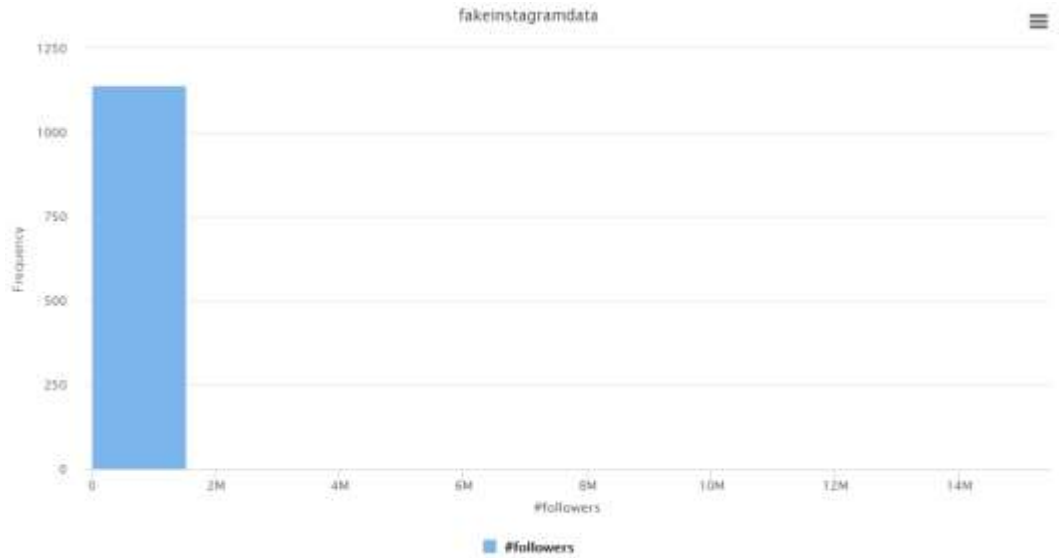
Followers adalah jumlah akun yang mengikuti akun profil tersebut. Banyak jumlah followers terkadang adalah parameter apakah akun tersebut akun asli atau palsu. Dan biasanya sedikitnya jumlah followers juga karna akun baru sehingga besar kemungkinan adalah akun palsu. Dalam dataset ini followers pada akun di tentukan dengan kriteria antara 0 sampai 15.338.538 sesuai dengan batas followers pada salah satu data dalam dataset. pada perhitungan nilai dibuat perbandingan dengan >98 dan juga <98 berdasarkan perbandingan pada pohon keputusan. Berikut adalah jumlah nilai pada atribut followers.

Tabel 4. 11. Nominal Values pada followers

| No | Rentan Nilai $><$ | Jumlah Nilai | Fake Yes | Fake No |
|----|-------------------|--------------|----------|---------|
| 1 | <98 | 486 | 452 | 34 |
| 2 | >98 | 662 | 128 | 538 |

Pada data table diatas dapat dilihat bahwa jumlah nilai pada rentan nilai <98 lebih kecil dengan dengan jumlah kasus 486 dengan perbandingan lebih

besar data yes dari pada data fake no pada akun asli dan palsu. Sebaliknya nilai >98 lebih besar dengan jumlah nilai sebesar 662 dengan fake yes 128 dan fake no 538. berikut adalah visualisasi pada atribut followers.



Gambar 4. 15. Visualisasasi Atribut followers

k. Follows

Follows adalah jumlah akun yang diikuti dalam akun profil Instagram. Dalam dataset ini follows pada akun di tentukan dengan kriteria antara 0 sampai 7.500 sesuai dengan batas follows pada salah satu data dalam dataset. pada perhitungan nilai dibuat perbandingan dengan >342 dan juga <342 berdasarkan perbandingan pada pohon keputusan. Berikut adalah jumlah nilai pada atribut follows.

Tabel 4. 12. Nominal Values pada follows

| No | Rentan Nilai >< | Jumlah Nilai | Fake Yes | Fake No |
|----|-----------------|--------------|----------|---------|
| 1 | <342 | 682 | 470 | 212 |
| 2 | >342 | 470 | 110 | 360 |

Pada data table diatas dapat dilihat bahwa jumlah nilai pada rentan nilai >342 lebih kecil dengan dengan jumlah kasus 470 dengan perbandingan