

Nama Kelompok: Pegasus

Dataset: Banking - Marketings Target

1. Descriptive Statistics

- a. Secara tipe data sudah sesuai, tetapi ada kolom yang nilainya mungkin butuh perhatian khusus.
Kolom balance yang seharusnya menampung nilai rerata saldo(balance) tahunan ada yang bernilai negatif dan mungkin ini disebabkan tanggungan seperti biaya administratif bank yang sudah menumpuk (hanya asumsi).
Kolom pdays memiliki kondisi khusus di mana nilai -1 berarti client belum pernah di-contact dan ini tentunya mempengaruhi nilai mean, percentile, serta standar deviasi keseluruhan. Ke depannya, pun seperti ini butuh penanganan khusus.
Pada kolom duration terdapat 3 baris data dengan nilai 0 (outcome = unknown) yang tipe contact-nya juga bertipe 'unknown', seperti ini bisa dihilangkan saja karena kemungkinan client tidak menerima panggilan dan validitasnya kurang dapat dipastikan.
Untuk kedepannya, kolom - kolom dengan nilai 'unknown' mungkin bisa diberikan penanganan khusus, terutama jika kolom tersebut banyak yang mengandung nilai 'unknown'
- b. Tidak ada kolom yang memiliki nilai kosong/null
- c. Kolom balance memiliki nilai min yang bernilai -8019 dan max sebesar 102127 dan jika dilihat dari percentile 75% bernilai 1428, data balance bersifat skewed positive (Mungkin bisa diterapkan rescaling). Kolom pdays memiliki nilai min, Q1, Q2, dan Q3 yang bernilai -1, ini berarti sebagian besar dari client belum dihubungi 'lagi' oleh pihak campaign marketing.

rekomendasi: Kolom month yang berisi nama bulan bisa saja diubah menjadi angka (int) yang mewakili bulan (jan -> 1, des -> 12) (Label encoding) Begitu juga untuk kolom default, housing, loan, dan y yang bertipe binary (yes/no) menjadi 0 dan 1.

2. Univariate Analysis

- a. Berdasarkan boxplot terdeteksi outlier hampir di seluruh kolom, yaitu pada kolom age, balance, duration, campaign, pdays, dan previous.
Kolom age terlihat cenderung skewed, dan kolom balance, duration, campaign, pdays, serta previous sangat jelas terlihat bahwa pola distribusinya skewed.
- b. Distplot menegaskan analisa boxplot, yaitu sebaran data hampir seluruhnya positively skewed, kecuali pada kolom day yang cenderung normal.
- c. Violin plot merupakan gabungan dari boxplot dan distplot, terlihat bahwa kolom age cenderung skewed. Kolom balance, duration, campaign, pdays, dan previous terlihat sangat skewed (positively). Kolom day cenderung normal.
- d. Berdasarkan hasil analisa boxplot, distplot, dan violin plot, didapatkan bahwa data numerical direkomendasikan untuk outlier handling, log/exp transformation, dan normalisasi/standardisasi.
- e. Berdasarkan hasil analisa count plot kolom job dan month menunjukkan kardinalitas (jumlah unique values) yang cukup tinggi.
Status pernikahan nasabah didominasi oleh yang telah menikah.

Mayoritas jenjang pendidikan adalah secondary, tidak memiliki credit in default, memiliki pinjaman rumah, dan memiliki pinjaman pribadi.
Mayoritas komunikasi dilakukan dengan seluler dan bulan terakhir kali nasabah dihubungi CS kebanyakan adalah mei.
Hasil campaign sebelumnya mayoritas tidak diketahui, diluar itu jumlah campaign yang gagal lebih banyak dari yang berhasil.
Hasil akhirnya, berdasarkan data yang ada, nasabah yang membuka rekening deposito jauh lebih sedikit daripada yang tidak membuka.
Kolom month dan job didrop saja karena jumlah unique value terlalu banyak.
Perlu encoding untuk kolom - kolom categorical, yaitu default, housing, loan, dan y menggunakan label encoding, marital, education, contact, dan poutcome menggunakan one hot encoding.

3. Multivariate Analysis

- a. Pada analisis menggunakan heatmap, tidak ditemukan fitur-fitur numeric yang berkorelasi cukup kuat dengan target (deposito). Sehingga tidak dapat ditarik kesimpulan fitur mana yang dapat dipertahankan menggunakan analisis heatmap. Pada analisis menggunakan pairplot terdapat beberapa kemungkinan korelasi antara target dengan beberapa fitur :
 - Pada hubungan antara age dengan pembukaan deposito, customer dengan umur yang lebih tinggi cenderung untuk tidak membuka deposito.
 - Customer dengan jumlah campaign yang banyak cenderung tidak membuka deposito.
 - Beberapa fitur juga menunjukkan adanya korelasi dengan keputusan membuka deposito antara lain previous & duration, namun belum pasti causalitasnya. Perlu direview ulang.

Pada analisis category column vs target variable : pada category housing persebaran customer yang memiliki pinjaman perumahan cenderung tidak membuka deposito, sedangkan customer yang tidak memiliki pinjaman perumahan cenderung membuka deposito.

Fitur yang dapat yang cukup relevan dan dapat dipertahankan antara lain : age, campaign, housing

- b. Pada analisis menggunakan heatmap, tidak ditemukan pola dan korelasi yg kuat (>0.7) antar fitur. Meski demikian, antara fitur duration - y (keputusan membuka deposito) dan fitur previous - pdays memiliki korelasi yang lumayan tinggi dan cukup menarik yang di rekomendasikan untuk coba dilanjutkan ke tahap *pre-processing*. Pada analisis menggunakan pairplot, dapat diasumsikan tidak ditemukan pola yang menarik antar fitur.

4. Business Insight

Berdasarkan hasil uji hipotesis dan analisis fitur yang ada didapatkan beberapa insight dan rekomendasi:

- a. Berdasarkan grafik hubungan antara persentase jumlah orang yang menerima/mengambil deposit dengan waktu(bulan) telepon marketing dilakukan didapatkan:
- Insight: persentase customer yang mengambil deposit tertinggi terjadi pada saat telepon marketing dilakukan di bulan maret
 - Rekomendasi yang bisa dilakukan ialah dengan memperbanyak telepon marketing pada bulan tersebut, karena pada bulan tersebut persentase customer untuk menerima tawaran deposit lebih tinggi dibandingkan bulan-bulan lainnya.
- Rekomendasi lain adalah dengan mengurangi jumlah telepon marketing pada bulan-bulan yang memiliki persentase customer untuk menerima tawaran yang rendah
- b. Berdasarkan grafik distribusi jumlah campaign untuk customer yang tidak maupun yang mengambil deposit didapatkan,
- Insight:Insightnya diambil berdasarkan hasil uji hipotesis, yaitu semakin tinggi jumlah campaign yang dilakukan kepada customer akan semakin berkurang/sedikit jumlah customer yang mengambil akun deposit maupun yang tidak mengambil akun deposit. Terlalu sering melakukan campaign(telepon marketing) terhadap customer, malah menurunkan keinginan user untuk mengambil mengambil deposit maupun tidak mengambil deposit
 - Rekomendasi: Jumlah campaign yang terlalu banyak tidak memberikan kepastian bahwa customer akan mengambil deposit ataupun tidak mengambil deposit. Perlu ditentukan jumlah campaign yang optimum, sehingga jumlah campaign(biaya marketing telepon) dapat berkurang dan customer memiliki kepastian yang lebih tinggi untuk mengambil deposit ataupun tidak mengambil deposit.
- c. Berdasarkan grafik distribusi durasi campaign untuk customer yang tidak maupun yang mengambil deposit didapatkan,
- Insight:Insightnya diambil berdasarkan hasil uji hipotesis, yaitu semakin lama durasi campaign yang dilakukan kepada customer akan semakin berkurang/sedikit jumlah customer yang mengambil akun deposit maupun yang tidak mengambil akun deposit. Terlalu lama durasi telepon yang dilakukan terhadap customer, malah menurunkan keinginan user untuk mengambil mengambil deposit maupun tidak mengambil deposit
 - Rekomendasi: Durasi campaign yang terlalu lama tidak memberikan kepastian bahwa customer akan mengambil deposit ataupun tidak mengambil deposit. Perlu ditentukan durasi campaign yang optimum, sehingga durasi campaign(biaya marketing telepon) dapat berkurang dan customer memiliki kepastian yang lebih tinggi untuk mengambil deposit ataupun tidak mengambil deposit.