|  |
| --- |

| No. | Topic/Subject | Action | Action by | Remark |
| --- | --- | --- | --- | --- |
| 1 | **Goals and Business Metrics**   * Dalam menentukan goals, sebaiknya ditetapkan 1 saja. Hal ini bertujuan agar tim bisa lebih fokus kepada goals tersebut. * Pastikan goals dapat diukur secara kuantitatif. * Goal untuk Banking Dataset: Meningkatkan conversion rate nasabah untuk membeli deposito. * Business metrics CvR secara tidak langsung mempengaruhi cost acquisition. Semakin tinggi CvR, maka akan semakin rendah cost acquisition. | Membuat goal menjadi 1 saja |  |  |
| 2 | **Exploratory Data Analysis**   * Untuk orang/stakeholder non-data, dengan melihat mean dan median sekilas dapat diketahui bentuk distribusi datanya. Ada metode khusus untuk melihat distribusi data yaitu uji normalitas. * Berdasarkan visualisasi tentang boxplot duration vs contact dengan hue target, dapat terlihat bahwa user dengan duration yang lebih tinggi cenderung untuk mendaftar deposito. Namun hal ini perlu dibuktikan lebih jauh dengan hipotesis karena belum ada bukti konkrit yang menjelaskan kausalitas duration terhadap target. Meskipun begitu, dapat dibuat suatu asumsi bahwa sebaiknya kita dapat melakukan teknik persuasi yang baik di 2-3 menit awal kontak agar user tertarik untuk mendengar lebih lama. * Feature default (gagal bayar) dan loan (utang) berkaitan dengan performa kredit user. * Ketika melakukan EDA dengan kolom categorical, sebaiknya dibuat juga dalam bentuk persentase. Membuat persentase setiap nilai dari suatu kolom dan pengaruhnya terhadap target dapat digunakan oleh tim bisnis dan marketing untuk menentukan customer yang lebih sesuai untuk diberi campaign. * Berdasarkan visualisasi histplot terkait range umur user, sebaiknya dibuat persentase setiap range umur untuk melihat pengaruhnya terhadap target. * Dari visualisasi awal, terlihat user dengan rentang umur 20 – 30 cenderung untuk subscribe ke deposito. Diasumsikan user dengan rentang umur ini belum memiliki loan sehingga masih cukup untuk menyimpan uangnya di deposito. * Perlu dibuat persentase pengaruh jumlah campaign untuk melihat pengaruhnya terhadap target. * Secara umum tahap Exploratory Data Analysis bertujuan untuk menggali insight-insight dan temuan menarik dari data. |  |  |  |
| 3 | **Data Pre-processing**   * Dilihat dari dataset yang banyak memiliki kolom imbalance, sepertinya perlu dilakukan oversampling/undersampling pada data. * Pada stage 2 – Data Preprocessing nanti, akan dibuat 3 skenario untuk mengolah data:  1. Diolah tanpa oversampling atau undersampling 2. Diolah dengan oversampling dan undersampling 3. Diolah dengan oversampling atau undersampling (pilih salah satu)  * Jumlah anggota tim Harta, Tahta, Data ada 8 orang. Untuk tahap selanjutnya mungkin dapat dibuat pre-processing data dengan 8 skenario pula. |  |  |  |