

UTS

1. Volume adalah dimana ukuran data pada big data sangat besar contohnya NoSQL DB, Velocity adalah kecepatan yang sangat cepat di mana data diterima dan (mungkin) langsung digunakan contohnya smart devices menggunakan data real time dalam pengoperasiannya, Variety adalah berbagai jenis data yang tersedia, ada yang terstruktur (sudah memiliki atribut eksplisit) dan tidak terstruktur (tidak memiliki atribut eksplisit) contohnya yang terstruktur yaitu data pegawai, dan contoh yang tidak terstruktur yaitu data yang berasal dari interface facebook dan lain lain.
2. Menurut saya yang lebih tepat menggunakan dictionary(Zunaidi) karena pada dictionary setiap penggunaan data berkelompok dapat kita tentukan sendiri elemennya dan indexnya sehingga memudahkan dapat memudahkan pengelolaan biodata. Dan sebenarnya list juga mampu melakukan demikian hanya saja index elemennya tidak dapat kita tentukan sendiri.
3. Baju zirah dan senjata yang tepat adalah baju zirah perak dan senjatanya kapak karena berdasar dari ke 10 data tersebut titania lebih condong mendekati ke andhika. Alasannya adalah tinggi badan titania 165 cm dan beratnya 53 kg, kemiripan tinggi badan yang paling mendekati adalah dengan tinggi badan andhika yakni 163 cm dan kemiripan berat badan yang paling dekat adalah dengan dini 55 kg, namun jika dihitung berdasar kecenderungan data yang didapat, titania 165 hanya lebih 2 cm dan berat badan selisih 4 kg dari andhika. Jika dicocokkan dengan dini, titania lebih tinggi 5 cm dan berat badan selisih 2 kg. Dan ketika dibandingkan dengan spek andhika $2 \text{ cm} + 4 \text{ kg} = 6 \text{ (cost/jarak)}$, sedang spek dini $5 \text{ cm} + 2 \text{ kg} = 7 \text{ (cost/jarak)}$, maka cost yang diambil adalah yang paling dekat/murah/sedikit. Maka yang lebih tepat menggunakan spek yang digunakan oleh andhika yakni perak dan kapak.

Langkah :

1. Menghitung selisih data titania dengan setiap data yang ada didalam table data.
 2. Menjumlahkan tb dan bb hasil perhitungan langkah 1
 3. Membandingkan hasil perhitungan langkah 2
 4. Mencari hasil yang paling murah/sedikit/dekat
 5. Menentukan jenis baju zirah dan senjata.
4. a) Menurut saya kasus missing value tersebut bersifat MCAR karena hal ini bersifat random events, bisa jadi si Martinus lupa mengkalkulasi dan lupa harga setiap bahan yang ia data

sehingga dia memilih untuk tidak mengisi harganya, atau bisa jadi juga si Martinus tidak tahu menahu (tidak melakukan sidak pasar/hal-hal lain yang bersifat random) tentang bahan masakan tersebut.

- b) Dengan menggunakan metode imputation teknik knn, langkahnya adalah dengan memberikan asumsi nilai/bobot (logika pasar, misal ayam > garam) pada setiap unit bahan (misal: Ayam=5, Garam=1 dst.) setelah itu melakukan penjumlahan nilai elemen pada atribut bahan pada masing-masing masakan. Setelah didapatkan nilai total dari setiap bahan-bahan yang digunakan pada setiap masakan setelah itu melakukan operasi knn, jika nilai total bahan masakan pada masakan ayam bakar bacem manis telah diketahui maka lakukan pencarian pada nilai total bahan masakan lain yang terdekat dengan nilai total yang dimiliki masakan ayam bakar bacem manis. Maka harga yang didapatkan (value yang missing) akan mengikuti pada nilai total masakan terdekat.