## **2.4 *Verify Data Quality***

Tahapan ini berisi evaluasi dan kualitas data dan kelengkapan data yang digunakan. Terjadinya *error* maupun kesalahan ketika *input* data mengakibatkan terjadinya *missing value* maupun *noise* pada data. Pada tahapan ini dilakukan pemeriksaan atribut yang hilang atau kosong. *Data cleaning* diperlukan untuk menjaga kekonsistenan dan menghilangkan data tidak relevan. *Data cleaning* pada proses *data mining* dapat mengurangi jumlah dan kompleksitas data. Memperkirakan apakah semua *value* dan ejaan nilai-nilai rasional serta apakah fitur dengan *value* yang berbeda memiliki pengertian yang sama. Hasil penelusuran yang dilakukan menemukan :

1. Terdapat atribut yang memiliki nilai *null* atau kosong, seperti pada atribut “education” dan “rating”, terdapat *cell* yang memiliki nilai *null* atau kosong.
2. Format tipe data pada atribut “rating” dalam bentuk integer dan “is\_promoted” yang dibuat dalam *binary* yang disajikan dalam angka akan menyebabakan kesulitan dalam memahami data untuk menemukan asosiasi.