**So sánh 2 bài ref (Telcom Customer Churn)**

Bài 1: https://towardsdatascience.com/churn-prediction-770d6cb582a5

Bài 2: https://towardsdatascience.com/churn-prediction-3a4a36c2129a

* Giống nhau:
* Đều có những bước tiền xử lý và đánh giá sự tác động của từng feature đối với target variable
* Khác nhau:

|  |  |  |
| --- | --- | --- |
|  | Bài 1 | Bài 2 |
| Encoding | - Categorial feature thành 0 và 1 cả biến nhị phân và multiple values  - nummerial feature được scale theo phương sai đơn vị | - Categorial feature thành 0 và 1 với dữ liệu nhị phân; get\_dummies() với multiple values.  - nummerial feature được phân cụm sử dụng K-means(tìm k bằng elbow method) |
| Model | Sử dụng Mô hình Decision Tree | Sử dụng Mô hình phân loại nhị phân với XGBoost |
|  |  |  |
|  |  |  |