

# Data Intake Report

Name: Data Science : Bank Marketing (Campaign)

Report date: Feb 21 2024

Internship Batch: LISUM28

Version:1.0

Data intake by:Minseok Kim

Data intake reviewer:Minseok Kim

Data storage location: <https://github.com/N0VA-code/Week12.git>

## Tabular data details:

Total number of observations	42780
Total number of files	1
Total number of features	19
Base format of the file	.csv
Size of the data	4.6MB

**Note:** Replicate same table with file name if you have more than one file.

## Proposed Approach:

- Cleaned outlier using IQR.
- Predicting and evaluation via comparing 3 models(Logistic Regression

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.metrics import accuracy_score

df = pd.read_csv('/content/bank_cleaned.csv')

labelencoder = LabelEncoder()
categorical_columns = df.select_dtypes(include=['object']).columns
for column in categorical_columns:
    df[column] = labelencoder.fit_transform(df[column])

X = df.drop('y', axis=1)
y = df['y']
```

```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

#Logistic Regression
lr_model = LogisticRegression(max_iter=1000)
rf_model = RandomForestClassifier()
gb_model = GradientBoostingClassifier()

lr_model.fit(X_train, y_train)
lr_pred = lr_model.predict(X_test)
lr_accuracy = accuracy_score(y_test, lr_pred)

#Random Forest
rf_model.fit(X_train, y_train)
rf_pred = rf_model.predict(X_test)
rf_accuracy = accuracy_score(y_test, rf_pred)

#Gradient Boosting
gb_model.fit(X_train, y_train)
gb_pred = gb_model.predict(X_test)
gb_accuracy = accuracy_score(y_test, gb_pred)

lr_accuracy, rf_accuracy, gb_accuracy

```

**and the result was**

**(0.8868630201028518, 0.9145628798503974, 0.9159654043945769)**

**Based on this result, Our team will use Gradient Boosting model to this case.**