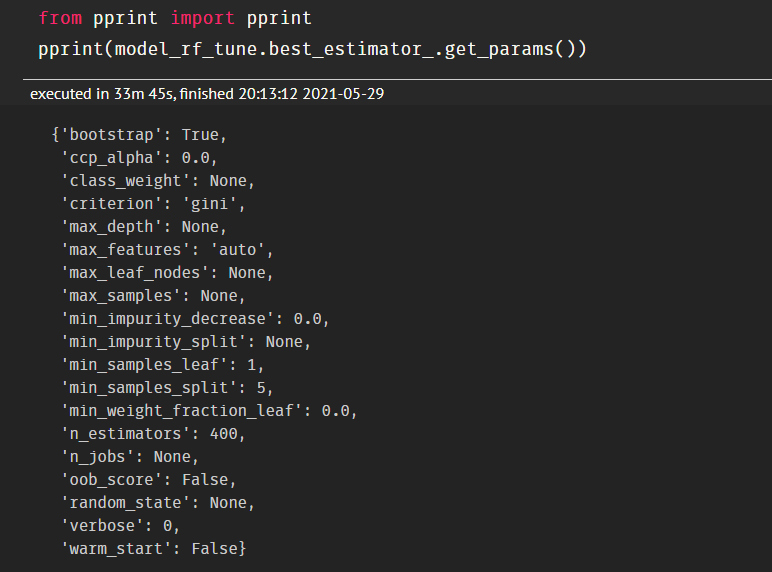
Approaches for Credict Card Lead predictions case:

**Approach on High level**

* Data pre-processing, analysis and model building.
* Tried various classical machine learning algorithms like Logistic Regression CV, Random Forest ,Gradient Boost and XGBOOST, in which Random Forest With tuned parameters worked best
* Random Forest with following parameters being used



**FEATURE ENGINEERING AND DATA PREPROCESSING:**

* Features has been selected on the basis of common business sense using excel.
* The are NaN values in the Credit\_Product feature in both train and test data. There is a good relationship of NaN compared with the target, created a new category in 'Credit\_Product' as No\_Info.
* From Feature Region\_Code removed the first two characters, as there is a high possibility that a particular region can some multi-baggers leads.
* Some outliers are seen for which a lead is present where age is more than 75+, but didn’t removed as there are less number of records for them and this can be a significant factor where older people are leads for their business requirement needs.
* 'Channel\_Code','Is\_Active','Occupation','Gender','Credit\_Product' columns are converted to dummies.
* ‘ID’ column is removed from the training data and similarly from the prediction data as well. Final results are combined with the respective ID.
* **Model Score:**

