Questions - Credit Default Risk

Many people struggle to get loans due to insufficient or non-existent credit histories. And, unfortunately, this population is often taken advantage of by untrustworthy lenders. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience, Home Credit makes use of a variety of alternative data--including telco and transactional information--to predict their clients' repayment abilities.

While Home Credit is currently using various statistical and machine learning methods to make these predictions, they're challenging Kagglers to help them unlock the full potential of their data. Doing so will ensure that clients capable of repayment are not rejected and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.

Data Dictionary:

| **Column Name** | **Description** |
| --- | --- |
| TARGET | Target variable (1 - client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample, 0 - all other cases) |
| NAME\_CONTRACT\_TYPE | Identification if loan is cash or revolving |
| CODE\_GENDER | Gender of the client |
| FLAG\_OWN\_CAR | Flag if the client owns a car |
| FLAG\_OWN\_REALTY | Flag if client owns a house or flat |
| CNT\_CHILDREN | Number of children the client has |
| AMT\_INCOME\_TOTAL | Income of the client |
| AMT\_CREDIT | Credit amount of the loan |
| AMT\_ANNUITY | Loan annuity |
| NAME\_INCOME\_TYPE | Clients income type (businessman, working, maternity leave,) |
| NAME\_EDUCATION\_TYPE | Level of highest education the client achieved |
| NAME\_FAMILY\_STATUS | Family status of the client |
| NAME\_HOUSING\_TYPE | What is the housing situation of the client (renting, living with parents, ...) |
| DAYS\_BIRTH | Client's age in days at the time of application |
| DAYS\_EMPLOYED | How many days before the application the person started current employment |
| OWN\_CAR\_AGE | Age of client's car |
| FLAG\_MOBIL | Did client provide mobile phone (1=YES, 0=NO) |
| FLAG\_EMP\_PHONE | Did client provide work phone (1=YES, 0=NO) |
| FLAG\_WORK\_PHONE | Did client provide home phone (1=YES, 0=NO) |
| FLAG\_CONT\_MOBILE | Was mobile phone reachable (1=YES, 0=NO) |
| FLAG\_PHONE | Did client provide home phone (1=YES, 0=NO) |
| OCCUPATION\_TYPE | What kind of occupation does the client have |
| CNT\_FAM\_MEMBERS | How many family members does client have |
| REGION\_RATING\_CLIENT | Our rating of the region where client lives (1,2,3) |
| REGION\_RATING\_CLIENT\_W\_CITY | Our rating of the region where client lives with taking city into account (1,2,3) |
| REG\_REGION\_NOT\_LIVE\_REGION | Flag if client's permanent address does not match contact address (1=different, 0=same, at region level) |
| REG\_REGION\_NOT\_WORK\_REGION | Flag if client's permanent address does not match work address (1=different, 0=same, at region level) |
| ORGANIZATION\_TYPE | Type of organization where client works |
| FLAG\_DOCUMENT\_2 | Did client provide document 2 |
| FLAG\_DOCUMENT\_3 | Did client provide document 3 |
| FLAG\_DOCUMENT\_4 | Did client provide document 4 |
| FLAG\_DOCUMENT\_5 | Did client provide document 5 |
| FLAG\_DOCUMENT\_6 | Did client provide document 6 |
| FLAG\_DOCUMENT\_7 | Did client provide document 7 |
| FLAG\_DOCUMENT\_8 | Did client provide document 8 |
| FLAG\_DOCUMENT\_9 | Did client provide document 9 |
| FLAG\_DOCUMENT\_10 | Did client provide document 10 |
| FLAG\_DOCUMENT\_11 | Did client provide document 11 |
| FLAG\_DOCUMENT\_12 | Did client provide document 12 |

Perform the following operations:

1. Consider the given data and build a logistic regression model using spark ml
2. Save the model on HDFS
3. Consider a customer with following information:-

Cash loans,F,Y,Y,1,171000.0,1560726.0,41301.0,Commercial associate,Higher education,Married,House / apartment,-13778,-3130,1,1,0,1,1,3.0,2,2,0,0,Business Entity Type 3,0,0,0,0,0,0,1,0,0,0,0

Create a Spark Streaming application that reads the data from Kafka producer application and shows the prediction in real time for such data.