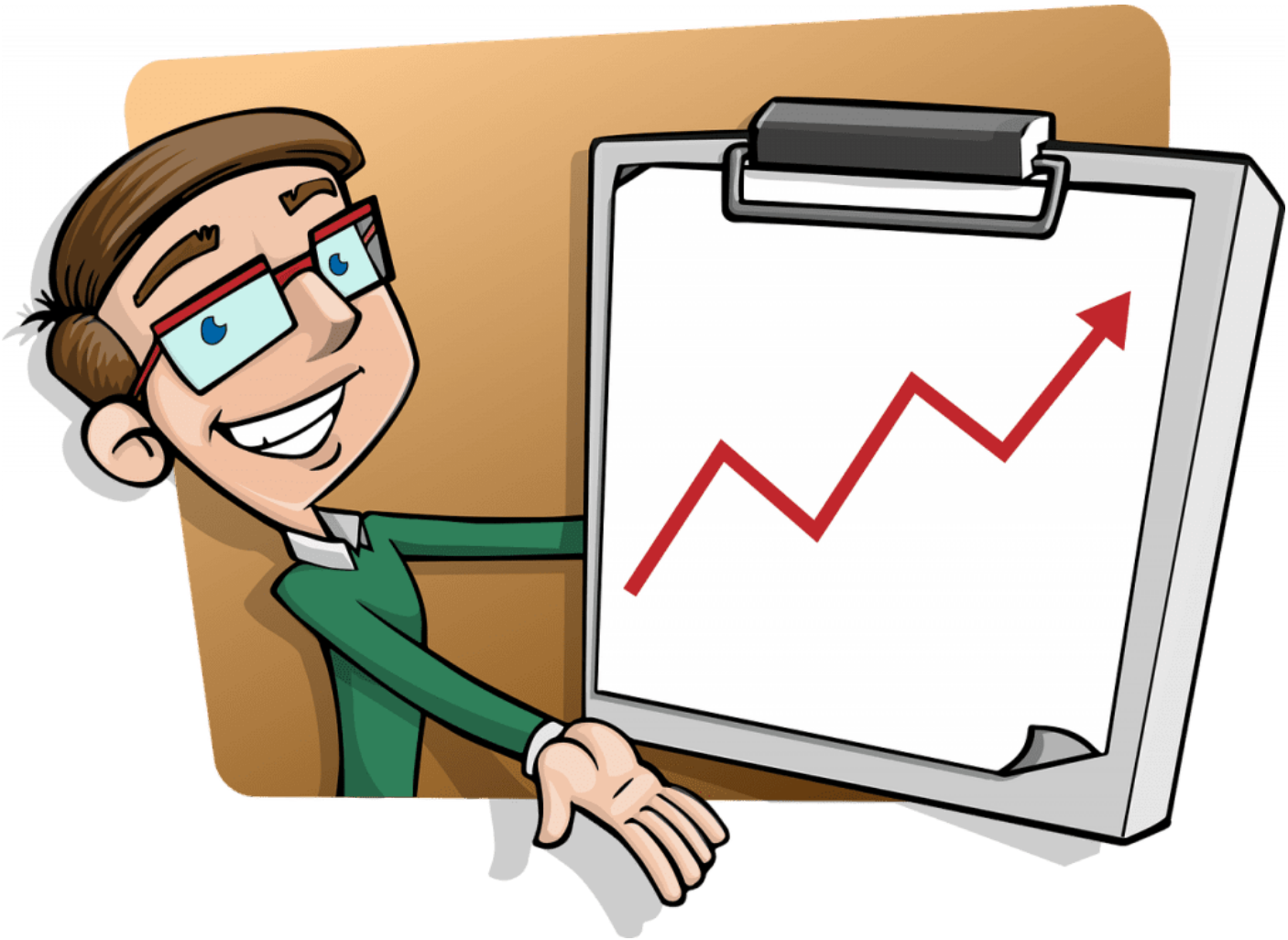

Machine Learning can Mitigate Risk for Lenders

23 minutes of reading

LendingClub is an American online credit marketplace that connects borrowers and investors. It offers credit solutions such as personal loans, business loans, and refinancing solutions to borrowers. The investors are banks and institutional investors who can filter and purchase loans through specific credit attributes that fit their investment and risk preferences. With more than 3 million members, LendingClub facilitates billions in loans each year.

income data. LendingClub then services the loan once investors make a lending decision.

While this business model saves investors money on marketing and loan origination cost, the return on investment is negatively impacted if the borrower defaults because most loans are unsecured – they have no collateral backing.

Business Understanding

An institutional investor/lender would like to purchase high-quality unsecured loans from LendingClub. Their goal is to create the right loan investment strategies and mix that minimizes risk and maximizes the overall returns from this loan portfolio. They want to predict the loan portfolio is profitable while pre-purchasing it.

> ***"So answer the question – who shall we buy?"***

An important consideration is that LendingClub does not require collateral for unsecured loans. Therefore, should a borrower default, the possibility of recovering the outstanding principal and interest is very low and the overall return of the portfolio is adversely impacted if such an account is eventually designated as charged-off. Therefore, reliably predicting if a loan will be fully-paid or charged-off would reduce credit risk and the associated financial loss for the lender.

- A loan is **charged-off** if the company believes it is unlikely to be collected as the borrower has substantially defaulted on payments. After a loan is charged off, the lender could sell the debt to a third-party collections agency that would attempt to collect on the delinquent account. A borrower owes the debt until it is paid off, settled, discharged in a bankruptcy proceeding, or in case of legal proceedings, becomes too old due to the statute of limitations.

Credit risk is the possibility of financial loss that may arise when a borrower fails to make the required payment on loans to a lender. Financial loss could include the loss of principal and interest, disruption in cash flo, and increased collection costs.

Lenders gauge the creditworthiness of potential borrows using the five C's of credit.

1. **Character**, also called the credit history, measures the borrowers' reputation and history of repaying debt. Factors include length of employment, credit score, liens, and judgment reports.
2. **Capacity** measures the ability of the borrower to repay the loan by comparing income against debts. Lower debt-to-income ratio is usually preferred.
3. **Conditions** consider the interest rate, loan term, and loan amount.
4. **Capital** measures the contribution that the borrower puts towards a potential investment. Examples are down payments for mortgages and auto financing. LendingClub does not evaluate loan applications based on this criterion.
5. **Collateral** are assets that borrowers use to secure a loan. They give the assurance that the lender can recover all or a portion of the loan in the event of default. Collateral is not required for most of LendingClub loan products.

LendingClub collects data such as employment length, annual income, FICO score, number of open accounts, public record bankruptcies, tax liens, inquiries in the last 6 months, number of satisfactory accounts, and other loan specific information from all loan applications.

Data Description

The raw data was 584,862 rows, 152 columns, and 23,116,937 missing values:

Column name	Description
acc_now_delinq	The number of accounts on which the

	months.
addr_state	The state provided by the borrower in the loan application
all_util	Balance to credit limit on all trades
annual_inc	The self-reported annual income provided by the borrower during registration.
annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
avg_cur_bal	Average current balance of all accounts
bc_open_to_buy	Total open to buy on revolving bank cards.
bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts.
chargeoff_within_12_mths	Number of charge-offs within 12 months
collection_recovery_fee	Post charge off collection fee
collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections

	borrower's credit file for the past 2 years
delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
desc	Loan description provided by the borrower
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the co-borrowers' combined self-reported monthly income
earliest_cr_line	The month the borrower's earliest reported credit line was opened
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
emp_title	The job title supplied by the Borrower when applying for the loan. The

fico_range_high	The upper boundary range the borrower’s FICO at loan origination belongs to.
fico_range_low	The lower boundary range the borrower’s FICO at loan origination belongs to.
funded_amnt	The total amount committed to that loan at that point in time
funded_amnt_inv	The total amount committed by investors for that loan at that point in time
grade	LC assigned loan grade
home_ownership	The home ownership status provided by the borrower during registration or obtained from the credit report. The values are RENT, OWN, MORTGAGE, and OTHER.
id	A unique LC assigned ID for the loan listing.
il_util	Ratio of total current balance to high credit/credit limit on all install acct
initial_list_status	The initial listing status of the loan. Possible values are – W, F
inq_fi	Number of personal finance inquiries

inq_last_6mths	The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
installment	The monthly payment owed by the borrower if the loan originates
int_rate	Interest Rate on the loan
issue_d	The month which the loan was funded
last_credit_pull_d	The most recent month LC pulled credit for this loan
last_fico_range_high	The upper boundary range the borrower’s last FICO pulled belongs to.
last_fico_range_low	The lower boundary range the borrower’s last FICO pulled belongs to.
last_pymnt_amnt	Last total payment amount received
last_pymnt_d	Last month payment was received
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
loan_status	**Current status of the loan, the target feature.**
max_bal_bc	Maximum current balance owed on all revolving accounts

mo_sin_old_il_acct	Months since oldest bank instalment account opened
mo_sin_old_rev_tl_op	Months since oldest revolving account opened
mo_sin_rcnt_rev_tl_op	Months since most recent revolving account opened
mo_sin_rcnt_tl	Months since most recent account opened
mort_acc	Number of mortgage accounts
mths_since_last_delinq	The number of months since the borrower's last delinquency
mths_since_last_major_derog	Months since most recent 90-day or worse rating
mths_since_last_record	The number of months since the last public record
mths_since_rcnt_il	Months since most recent instalment accounts opened
mths_since_recent_bc	Months since the most recent bank card account opened
mths_since_recent_bc_dlq	Months since most recent bank card delinquency
mths_since_recent_inq	Months since the most recent inquiry
mths_since_recent_revol_delinq	Months since the most recent revolving

num_accts_ever_120_pd	Number of accounts ever 120 or more days past due
num_actv_bc_tl	Number of currently active bankcard accounts
num_actv_rev_tl	Number of currently active revolving trades
num_bc_sats	Number of satisfactory bankcard accounts
num_bc_tl	Number of bankcard accounts
num_il_tl	Number of installment accounts
num_op_rev_tl	Number of open revolving accounts
num_rev_accts	Number of revolving accounts
num_rev_tl_bal_gt_0	Number of revolving trades with balance >0
num_sats	Number of satisfactory accounts
num_tl_120dpd_2m	Number of accounts currently 120 days past due (updated in past 2 months)
num_tl_30dpd	Number of accounts currently 30 days past due (updated in past 2 months)
num_tl_90g_dpd_24m	Number of accounts 90 or more days past due in last 24 months
num_tl_op_past_12m	Number of accounts opened in past 12

	borrower's credit file.
open_acc_6m	Number of open trades in last 6 months
open_il_12m	Number of instalment accounts opened in past 12 months
open_il_24m	Number of instalment accounts opened in past 24 months
open_act_il	Number of currently active instalment trades
open_rv_12m	Number of revolving trades opened in past 12 months
open_rv_24m	Number of revolving trades opened in past 24 months
out_prncp	Remaining outstanding principal for total amount funded
out_prncp_inv	Remaining outstanding principal for portion of total amount funded by investors
pct_tl_nvr_dlq	Percent of trades never delinquent
percent_bc_gt_75	Percentage of all bank card accounts > 75% of limit.
policy_code	Publicly available policy_code=1 New products not publicly available policy_code=2

purpose	A category provided by the borrower for the loan request
pymnt_plan	Indicates if a payment plan is put in place for the loan
recoveries	Post charge off gross recovery
revol_bal	Total credit revolving balance
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit
sub_grade	LC assigned loan subgrade
tax_liens	Number of tax liens
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
title	The loan title provided by the borrower
tot_coll_amt	Total collection amounts ever owed
tot_cur_bal	Total current balance of all accounts
tot_hi_cred_lim	Total high credit/credit limit
total_acc	The total number of credit lines currently in the borrower's credit file
total_bal_ex_mort	Total credit balance excluding mortgage

total_bc_limit	Total bankcard high credit/credit limit
total_cu_tl	Number of finance trades
total_il_high_credit_limit	Total instalment high credit/credit limit
total_pymnt	Payments received to date for total amount funded
total_pymnt_inv	Payments received to date for portion of total amount funded by investors
total_rec_int	Interest received to date
total_rec_late_fee	Late fees received to date
total_rec_prncp	Principal received to date
total_rev_hi_lim	Total revolving high credit/credit limit
url	URL for the LC page with listing data.
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified
verified_status_joint	Indicates if the co-borrowers' joint income was verified by LC, not verified, or if the income source was verified
zip_code	The first 3 numbers of the zip code provided by the borrower in the loan application.
revol_bal_joint	Sum of revolving credit balance of the

	applicant
sec_app_fico_range_high	FICO range (low) for the secondary applicant
sec_app_earliest_cr_line	Earliest credit line at time of application for the secondary applicant
sec_app_inq_last_6mths	Credit inquiries in the last 6 months at time of application for the secondary applicant
sec_app_mort_acc	Number of mortgage accounts at time of application for the secondary applicant
sec_app_open_acc	Number of open trades at time of application for the secondary applicant
sec_app_revol_util	Ratio of total current balance to high credit/credit limit for all revolving accounts
sec_app_open_act_il	Number of currently active instalment trades at time of application for the secondary applicant
sec_app_num_rev_accts	Number of revolving accounts at time of application for the secondary applicant
sec_app_chargeoff_within_12_mths	Number of charge-offs within last 12 months at time of application for the secondary applicant

	time of application for the secondary applicant
sec_app_mths_since_last_major_derog	Months since most recent 90-day or worse rating at time of application for the secondary applicant
hardship_flag	Flags if the borrower is on a hardship plan
hardship_type	Describes the hardship plan offering
hardship_reason	Describes the reason the hardship plan was offered
hardship_status	Describes if the hardship plan is active, pending, canceled, completed, or broken
deferral_term	Amount of months that the borrower is expected to pay less than the contractual monthly payment amount due to a hardship plan
hardship_amount	The interest payment that the borrower has committed to make each month while they are on a hardship plan
hardship_start_date	The start date of the hardship plan period
hardship_end_date	The end date of the hardship plan period
payment_plan_start_date	The day the first hardship plan payment is due. For example, if a borrower has a hardship plan period of 3 months, the

	to make interest-only payments.
hardship_length	The number of months the borrower will make smaller payments than normally obligated due to a hardship plan
hardship_dpd	Account days past due as of the hardship plan start date
hardship_loan_status	Loan Status as of the hardship plan start date
orig_projected_additional_accrued_interest	Original projected additional interest amount that will accrue for the given hardship payment plan as of the Hardship Start Date. It will be null if the borrower has broken their hardship payment plan.
hardship_payoff_balance_amount	The payoff balance amount as of the hardship plan start date
hardship_last_payment_amount	The last payment amount as of the hardship plan start date
disbursement_method	The method by which the borrower receives their loan. Possible values are: CASH, DIRECT_PAY
debt_settlement_flag	Flags whether or not the borrower, who has charged-off, is working with a debt-settlement company.
debt_settlement_flag_date	The most recent date that the

	plan. Possible values are: COMPLETE, ACTIVE, BROKEN, CANCELLED, DENIED, DRAFT
settlement_date	The date that the borrower agrees to the settlement plan
settlement_amount	The loan amount that the borrower has agreed to settle for
settlement_percentage	The settlement amount as a percentage of the payoff balance amount on the loan
settlement_term	The number of months that the borrower will be on the settlement plan

To understand the data better, features such as ‘employment length, annual income, FICO score, number of open accounts, public record bankruptcies, tax liens, inquiries in the last 6 months, and number of satisfactory accounts’ were plotted against the feature ‘Loan Status’ to provide a visual representation of the relationship that exists between the dependent (target) and independent variables.

It was observed that borrowers with 10+ years length of employment are most likely to fully pay their loans. Other employment lengths also exhibit a relationship. Other features showed a similar relationship to ‘Loan Status’.

Another analytic was to plot the ‘Purpose’ of the loan against the ‘Loan Status’.

It shows that debt consolidation loans are most likely to be paid off. A similar relationship was observed for all purposes.

The ‘loan status in percentage’ diagram shows that 15.70% of the loans were charged off while 84.30% were fully paid. This was a good observation as we will have to balance this for the purposes of machine learning later.

A model that can reliably predict whether a loan will be fully paid or charged-off could help reduce the financial loss to the investor's credit portfolio. The 'Loan_Status' is the target variable. Other features, also called dependent variables, will be used to predict the Loan_Status.

A ML model was created to predict the likelihood that a borrower who passes the initial screening but may default on the loan subsequently. The dependent variable (also called target feature) was 'Loan_Status'. It had two categories, fully-paid or charged-off.

Braintoy mlOS was used to wrangle the data and create the ML models. In conjugation, Tableau was used for data visualization.

Data Engineering

Data Wrangling

First, columns with less relevant data were deleted. Examples are data that cannot be estimated/obtained before the loan is disbursed, columns with the same values, duplicate data, or columns that show a similar relationship to Loan_Status as other columns.

List of Deleted Features	
id	acc_open_past_24mths
funded_amnt	avg_cur_bal
funded_amnt_inv	bc_open_to_buy
installment	bc_util
grade	chargeoff_within_12_mths
emp_title	delinq_amnt
home_ownership	mo_sin_old_il_acct
loan_status	mo_sin_old_rev_tl_op
pymnt_plan	mo_sin_rcnt_rev_tl_op
url	mo_sin_rcnt_tl
title	mort_acc
zip_code	mths_since_recent_bc
addr_state	mths_since_recent_bc_dlq

mths_since_last_record	mths_since_recent_revol_delinq
revol_bal	num_tl_120dpd_2m
revol_util	num_tl_30dpd
out_prncp	num_tl_90g_dpd_24m
out_prncp_inv	pct_tl_nvr_dlq
total_pymnt	percent_bc_gt_75
total_pymnt_inv	tot_hi_cred_lim
total_rec_prncp	total_bal_ex_mort
total_rec_int	total_bc_limit
total_rec_late_fee	total_il_high_credit_limit
recoveries	revol_bal_joint
collection_recovery_fee	sec_app_fico_range_low
last_pymnt_d	sec_app_fico_range_high
last_pymnt_amnt	sec_app_earliest_cr_line
next_pymnt_d	sec_app_inq_last_6mths
last_credit_pull_d	sec_app_mort_acc
policy_code	sec_app_open_acc

dti_joint	sec_app_open_act_il
verification_status_joint	sec_app_num_rev_accts
acc_now_delinq	sec_app_chargeoff_within_12_mths
tot_coll_amt	sec_app_collections_12_mths_ex_med
tot_cur_bal	hardship_flag
open_acc_6m	hardship_type
open_act_il	hardship_reason
open_il_12m	hardship_status
open_il_24m	deferral_term
mths_since_rcnt_il	hardship_amount
total_bal_il	hardship_start_date
il_util	hardship_end_date
open_rv_12m	payment_plan_start_date
open_rv_24m	hardship_length
max_bal_bc	hardship_dpd
all_util	hardship_loan_status
total_rev_hi_lim	orig_projected_additional_accrued_interest

total_bal_il	hardship_last_payment_amount
inq_last_12m	debt_settlement_flag

Second, all rows with the value of ‘Joint’ on the ‘Application type’ column and all columns with information about joint applications were deleted because borrowers who struggle to qualify for loans on their own may usually add a co-borrower to qualify and obtain favorable terms. Including these records may lead to a low-quality model. Therefore, it was important to exclude joint applications and rather have a separate analysis for that portion of the dataset.

Third, substrings “%” and “months” were deleted from interest rate and term respectively and the data type converted to numeric. At this stage, there were 39 columns and 452,069 samples.

After these data wrangling operations, it was observed that the ‘loan status’ column (the dependent/target variable) had 90,817 records for charged-off and 361,252 records for fully-paid loans.

Because the number of fully-paid loans far outweigh the charged-off loans, training a machine learning model on this dataset will lead to a biased model.

There are several techniques to resolve an imbalanced data problem such as creating synthetic data for the minority class, but because the sample size was large, random resampling (undersampling) to delete 97% of rows where the value equals ‘Fully paid’ and 90% of rows where the value equals “Charged Off” was applied on the loan status column.

This now became a balanced dataset that is fit for machine learning.

Feature Selection

This is the process of selecting the input features and the target feature and finding the relative importance of such features to predict the ‘Loan Status’.

Feature importance is a technique that is used to identify crucial features for an efficient model. Features with higher scores will have a higher effect on the model than features with a lower score. In addition to the benefit listed above, feature selection saves time and money by eliminating irrelevant features with lower scores.

The table below shows the score for the top ten features. In addition to these top ten features, the employment length, annual income, loan purpose, and initial list status were also selected to train the ML model.

S/N	Input feature	Importance score
1	last_fico_range_high	0.30
2	last_fico_range_low	0.27
3	int_rate	0.06
4	sub_grade	0.04
5	verification_status	0.04
6	loan_amnt	0.02
7	dti	0.01

10	open_acc	0.01
11	initial_list_status	0.01
12	emp_length	
13	annual_inc	
14	purpose	

Feature Preprocessing

This is the process of converting raw data into a specific format to make it work well for machine learning.

Categorical to numerical refactoring were applied on columns with categorical values such as sub_grade, emp_length, verification_status, loan_status, purpose, initial_list_status, and application_type.

training dataset is used to train the ML models and the validation dataset is used to evaluate the model's performance.

For this analysis, 80% of the dataset was used to train the model while 20% was used for validation.

Model Building

The purpose of the model is to predict if a loan will be fully paid or charged off. This is a binary decision – yes or no. Hence, it is a classification problem.

Classification is a supervised approach in machine learning where the algorithms learn from the provided data input and can then classify new data from what they learnt.

Automated machine learning was performed that created several models using various classification algorithms.

Of all the algorithms, the Extra Tree Classifier provided the highest accuracy of 95.41%. It was selected for publishing.

There were several other challenger models but the top three are Random Forest Classifier with accuracies of 95.35%, Quadratic Discriminant Analysis with an accuracy of 94.34%, and XGB classifier with an accuracy of 94.21%.

Model Evaluation

Confusion Matrix

Confusion matrix is a table that visualizes the performance of a model by showing the number of False Positives, False Negatives, True Positives, and True Negatives. A good matrix has larger numbers on the diagonal with a darker shade of blue and small numbers on the lighter shade of blue.

loans were correctly predicted as paid-off. Only 79 (2.2%) were incorrectly predicted as charged-off, and only 89 (2.4%) were incorrectly predicted as fully paid. Therefore, the percentage of loans that will eventually be charged off can be reduced from 15.7% (from the ‘Loan status in percentage’ diagram) to 2.4% by what the model learnt.

ROC Curve

The Receiver Operating Characteristic (ROC) curve is used to show how well a model can distinguish between True Values and Predicted Values. This is an indication of if the model has truly learned or simply memorized from the training data.

The more the curve fits the top left corner of the plot, the better the classification process.

This diagram shows that the Area Under the Curve (AUC) is 0.99, and it closely hugs the top left corner of the plot.

This is a good model!

Model Governance

A reviewer has to now double-check the published model/(s) for biases, inefficiencies, and inaccuracies. Model/(s) may be accepted as-is or may be rejected with a request to the modeler to improve it and publish a better version for review.

This maker-checker method brings prudence to machine learning operations (MLOps). Automated documentation makes it easy.

Jaspreet Gill of Braintoy was the coach for this internship. She reviewed the published model.

To start with, the Training and the Test sample were checked for ‘skewness’.

train the model vs. the data to test it were dissimilar? The performance metrics will have lesser meaning then. The ‘skew’ gives a visual hint that there might be something wrong.

The performance scores could be relied on if the distribution of data is similar for each feature in the training and test set.

It was observed that the distribution of training vs. test for each feature were similar. In addition, the target variable was balanced. The performance metrics could be relied on.

This was a supervised classification problem that shows an accuracy of 95.41%. This seemed to be a good model. But while accuracy is the simplest measure to evaluate model performance, the other evaluation criteria such as ROC curve, Confusion Matrix, F1-score, Hamming Loss, Precision, Recall, and the Jaccard Score are also important.

- **Precision** answers the question – what proportion of positive identifications were actually correct? **Recall** answers the question – What proportion of actual positives were correctly classified? Both Precision and Recall scores were 0.95 indicating that the model is performing well.
- The **F1** score is the harmonic mean of Precision and Recall, a summary indicator. A score of 95.41 indicated good performance.
- The **Hamming Loss** indicates the fraction of records that were incorrectly predicted. While for binary classification problems such as this, it is just 1 minus accuracy, but it becomes more relevant for multi-class classifications as hamming loss averages the 'loss' from each class across the dataset. A score of 0.05 shows that the published model has less 'loss'.
- The **Jaccard Score** is a measure of similarity between predictions and the test set and their intersection and union. Measured between 0 and 1, higher score indicates that predictions overlap more with the test set. In this case, a score of 0.91 indicated good performance.

Based on these performance metrics as well as the **ROC** and the **Confusion Matrix** described in the previous section, the model was approved for production use.

Model Deployment

After the model is reviewed and approved, the next step is to integrate it into a production environment for practical business decision-making.

The first step is to create an app wherein the model will be containerized.

The app was created and named LendingClubLoanApp.

The second step is to select the newly created app under Deployable Apps.

In this step, the LendingClubLoanApp and the version v.2-v.934 of the model were selected, the version that was approved in the Model Governance step.

The third step is to containerize the model in the app.

There are two options – the app can be downloaded as a .zip file to a local computer or deployed to the cloud as an API/Microservice.

Option 1: Download app to local computer

Select “Download App” on the model deployment page.

Select HTML/Javascript as the code type, input the file name, and confirm the download.

The file was saved as “LendingClubLoanApp” on the local computer.

It could now be opened on any installed browser.

For security, the API keys and access token are required to run the app. Note that an ‘unauthorized’ error message is displayed in the model response log if the correct API key and access token are not entered.

To generate the API key and access token, the LendingClubLoanApp and “Manage Apps” buttons on the Model Deployment page were selected.

Then on the App Manager page, the ‘API Keys’ and ‘+ Add API KEY’ were selected.

Client ID and Client Secret were created on the API Key and Access Token page.

The public key and secret key were generated and saved as key pairs.

On the App Manager page, the saved Key Pair must be shared with the user for validation. It is only activated when shared.

The Key Pair was shared with myself to test the AI application.

Clicking on the View icon could show the API keys and Access Token.

The keys were copied to the appropriate sections on the HTML/Javascript page previously downloaded to the local computer.

The model is now ready for use.

To test the app, some known values were put to check that the model predicts that the loan will be paid.

- loan_amnt: 5000
- int_rate: 20.0
- sub_grade: D2
- emp_length: 1 year
- annual_inc: 42000.0
- verification_status: Not Verified
- purpose: other
- dti: 6.46
- inq_last_6mths: 0
- mths_since_last_delinq: 21
- open_acc: 4
- initial_list_status: w
- last_fico_range_high: 669
- last_fico_range_low: 665

As a second test, alternate values were put to check that the model predicts that the loan will be charged-off.

- loan_amnt: 5000
- int_rate: 20.0
- sub_grade: D2
- emp_length: 1 year
- annual_inc: 42000.0
- verification_status: Not Verified
- purpose: other
- dti: 46.46
- inq_last_6mths: 0
- mths_since_last_delinq: 21
- open_acc: 4

- last_fico_range_low: 445

Option 2: Deploy the app on the cloud

In this option, the app will run on the mlOS cluster and can be made accessible to anyone in the world with whom the API Keys are shared.

To deploy the model to the mlOS cluster, ‘Deploy’ was selected under model deployment and mlOS-mlCluster-01 as the destination type.

Once deployed, the status of the LendingClubLoanApp changed to “Running” which tells that the app is ready for use.

Dashboard

Select the app from the My Apps page to open it.

Realtime Prediction

It accepts one set of inputs at a time and predicts an output.

When one set of test values were input, the output was predicted that the loan would be fully paid.

For another set of test inputs, the app predicted an output that the loan will be charged off.

bulk for scoring and prediction.

In data scoring, the probability of each output class is calculated and ranked, and the output class with the highest probability is selected.

The file for the validation set was selected and scored.

As seen in the first row in the above screenshot, the probability of the loan being fully paid is 0.03 while the probability of the loan being charged off is 0.97. Hence, the loan is predicted to be charged off.

All rows in the dataset were scored and appropriate predictions were observed.

Summary

Predicting loans that will likely be charged off reduces financial losses to the portfolio mix of an investor. It was observed that the ML models could reduce the exposure to risky loans from 15.4% to 2.4%.

> ***"Get 13% more from your investment by buying loans that predict Fully Paid and avoiding loans that predict Charged Off."***

The business environment is dynamic. Situations change. If the variables used to build the ML model change, the effectiveness of the model also reduces causing model decay. It is prudent to periodically review changes to the data characteristics and keep the model current with improved versions.

Author

Tolu Alade is a finance professional with almost a decade of experience in compliance, commercial, and retail banking.

She has an MBA in Finance, a Bachelor's degree in Economics, and is a certified full stack developer with proficiency in Python. She has expertise in data analytics, machine learning, and data science techniques.

← Previous Post

Search …

Recent Posts

Machine Learning can Mitigate Risk for Lenders

Autocategorizing transactions

Education is better with AI

Combating Fake News with AI

Reliable Shipping Time Prediction

Truckers Win by Predicting Empty Miles

Categories

Agriculture (1)

Digital Transformation (11)

Education (1)

Financial Services (4)

Healthcare (1)

How To (17)

Oil and Gas (2)

Retail and Consumer (4)

Software and Internet (3)

Supply Chain Management (3)

Training

Machine Learning Basics

Build Your First Model

Email

GO

PO Box 47126, Creekside, Calgary AB T3P 0B9