Combatting fraudulent fintech with machine learning

Jonathan Fu[1] Mrinal Mishra[2]

October 20, 2021

[1]Center for Sustainable Finance and Private Wealth
Department of Banking and Finance
University of Zürich
jonathan.fu@bf.uzh.ch

[2]The Swiss Finance Institute
Department of Banking and Finance
University of Zürich
mrinal.mishra@bf.uzh.ch

The project in a nutshell

1. There have been increased concerns over problematic finance mobile apps in recent years
2. Efforts to combat "scam apps" limited by various issues:
    - Lack understanding of scale and scope of problem in app stores
    - Reactive rather than proactive solutions
    - "Whack-a-mole" approach of "scam app" developers
3. We leverage varied app data to systematize evidence and classify problematic apps
4. We complement existing solutions by applying ML techniques to **increase efficiency and speed of ex-ante vetting and ex-post monitoring**
5. Preliminary results are promising

Today's agenda

1. The problem
    - Framing the problem
    - Systematizing evidence with new data
2. The proposed solution
    - Methodology
    - Results
    - Model evaluation
3. Next Steps & Conclusion
4. Q & A / Feedback

The problem

Anecdotal Evidence & Media Coverage

There has been increased media coverage highlighting problematic finance mobile applications. Examples of recent sources include:

The Economist

Forbes

The New York Times

LET TRUTH PREVAIL
THE TIMES OF INDIA

QUARTZ AFRICA

MANILA BULLETIN

Past & current solutions

Stakeholders have started documenting issues and developing solutions:

1. **Supervisory / Regulatory responses**
   - Creation of / improvement of complaints channels
   - Registering digital lenders and platforms (e.g., PH-SEC, OJK)
   - Ensuring data privacy regulation (e.g., RBI, PH-NPC)
   - Expanding role of associations and codes of conduct (RBI-DLAI, OJK-AFPI, PH- SEC-Fintech Alliance)
   - Cautioning the public (e.g., RBI, PH-NPC)
   - Direct action to remove problematic apps (e.g., RBI)

2. **External stakeholders**
   - Monitoring social media (CGAP/World Bank)
   - RBI Working Group developing strategies

Limitations of current knowledge and solutions

1. Lack systematic understanding of scale and scope of problem amongst mobile finance apps
2. Reactive rather than proactive solutions
3. “Whack-a-mole” approach of “scam app” developers
4. Based on our recent review, still large number of problematic apps and users falling victim

Our proposed solution

- Leverage existing high-frequency app data and apply machine learning techniques to both static and real-time data
- Complement current solutions by offering method that can *improve both ex-ante vetting and ex-post monitoring*.

Data

1. **Primary data source:** Third-party app intelligence provider

2. **Types of data obtained / used:**
    - App meta-data
    - App review data
    - Historical download data

3. **Data coverage:** all “finance” category apps in Google Play store for 63 countries

4. **Time coverage:**
    - App meta-data: monthly from January 2020 - April 2021; does not include if *unpublished* prior to 2020
    - App review data: historical, covering until April 2021
    - App download data: historical, covering until April 2021

Example of app meta data (from legitimate provider)

Tala - Instant Loans

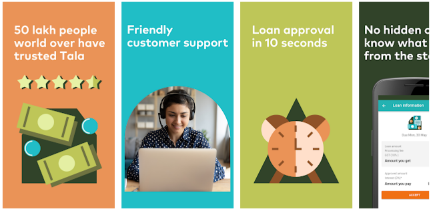
Tala Mobile Finance

5,705

PEGI 3

You don't have any devices

Add to Wishlist



Join lakhs of satisfied Tala customers who have managed expenses, paid for school fees, and grown their businesses with peace of mind.

How do I get a loan?

- Download the app and fill a quick form in the app
- Get your loan offer approval in under 10 seconds
- Verify your identity through our secure system
- Get your loan directly in your bank account
- No salary slip, bank statement or credit score needed.

What is the loan amount?

Your first loan amount starts from ₹500 - ₹1000. Grow with Tala and build your limit up to ₹10,000 with every on-time payment. Each loan has a tenure of 60 days.

How do I repay my loan?

- Make easy electronic payment in the app
- Choose to pay with debit card, UPI, netbanking, or wallets.
- Easy payment schedule keeps you on track with your payments

Get started. Download the Tala app now.

Tala accepts applications on behalf of its partners DMI Finance Private Limited and Apollo Finvest India Limited, licensed NBFCs regulated by RBI, to provide loans to its customers.

Updated	Size	Installs
September 27, 2021	13M	100,000+
Current Version	**Requires Android**	**Content Rating**
1.29.0	5.0 and up	PEGI 3 Learn more
Permissions	**Report**	**Offered By**
View details	Flag as inappropriate	Tala Mobile

Developer

hello@tala.co.in

Privacy Policy

1633 26th St. 3rd Floor

Santa Monica, CA 90404

Example of app meta data (from legitimate provider)

Tala - Instant Loans

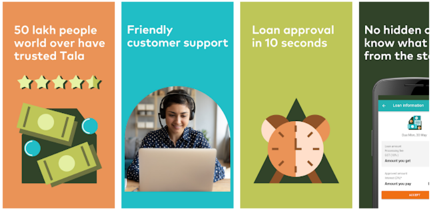
Tala Mobile Finance

★★★☆☆ 5,705

PEGI 3

You don't have any devices

Add to Wishlist



Join lakhs of satisfied Tala customers who have managed expenses, paid for school fees, and grown their businesses with peace of mind.

How do I get a loan?

- Download the app and fill a quick form in the app
- Get your loan offer approval in under 10 seconds
- Verify your identity through our secure system
- Get your loan directly in your bank account
- No salary slip, bank statement or credit score needed.

What is the loan amount?

Your first loan amount starts from ₹500 - ₹1000. Grow with Tala and build your limit up to ₹10,000 with every on-time payment. Each loan has a tenure of 60 days.

How do I repay my loan?

- Make easy electronic payment in the app
- Choose to pay with debit card, UPI, netbanking, or wallets.
- Easy payment schedule keeps you on track with your payments

Get started. Download the Tala app now.

Tala accepts applications on behalf of its partners DMI Finance Private Limited and Apollo Finvest India Limited, licensed NBFCs regulated by RBI, to provide loans to its customers.

Updated
September 27, 2021

Size
13M

Installs
100,000+

Current Version
1.29.0

Requires Android
5.0 and up

Content Rating
PEGI 3
Learn more

Permissions
View details

Report
Flag as inappropriate

Offered By
Tala Mobile

Developer
hello@tala.co.in
Privacy Policy
1633 26th St. 3rd Floor
Santa Monica, CA 90404

Meta data variables

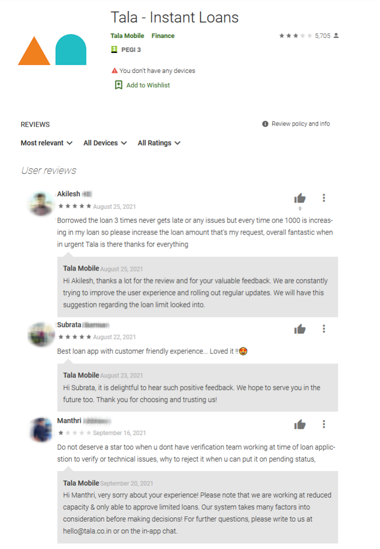
Visible data

Title
Short/long text description
Screenshots
Date last updated
Byte size
External version number
App developer
Developer website
Developer email
Developer physical address
Download / installs estimate (bucket)
# of ratings
# of reviews
Promo video

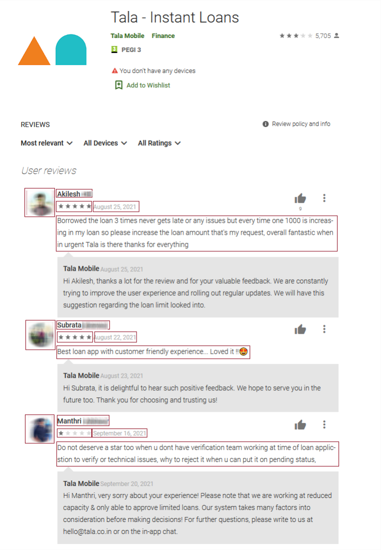
Hidden data

Date created
Disaggregated # ratings (1-5)
Internal version code
Software Development Kit (SDK) list
Permissions list
Whether contains ads
Number of reviews
App price
In-app purchases
Min/Max price for in-app purchases
Interactive elements
Countries available in
Main language
Languages available in
Other app-stores available in
Downloads (country-day level)

Example of review data (from legitimate provider)



Example of review data (from legitimate provider)



Review data

Rating (1 to 5)
Date of review
Review text
User name*
Profile pic*

**Not included in 3rd-party data provider's data but manually collecting*

The scale and scope of the finance app market (1/2)

To make pilot tractable, we categorize and narrow down to a subset of pure-play personal loan apps of interest:

1. Full sample: 134,744 “finance” category apps

2. We categorize apps broadly by product and provider type
   - Currently: “regular expressions” on meta-data titles, short, and long descriptions to tag product and provider types
     - *E.g., to tag personal loans: parse texts for “personal loan(s)”, “consumer loan(s)”, “payday loan(s)”, etc.*
   - Apply combinations of tags to filter to more precise subcategories

3. Current limitation
   - Note: About 65% of apps have English as main language (e.g., meta data descriptions)
   - Our analysis thus currently overlooks about one-third of finance-related apps

The scale and scope of the finance app market (2/2)

Table: Categorization of finance apps in sample (Jan 2020-April 2021)

	# of apps	% of apps	Est. # downloads (in M)	Avg. downloads per app
Panel A. Finance apps by EN vs. non-EN main language				
All finance	134,744	100.0%	6,970M	51,728
- **All finance (EN main lang.)**	**87,426**	**64.9%**	**4,080M**	**46,668**
- All finance (non-EN main lang.)	47,318	35.1%	2,890M	61,076
Panel B. Finance apps by product categories (EN only)				
All finance (EN main lang.)	**87,426**	–	**4,080M**	**46,668**
- General banking	18,015	20.6%	1,640M	91,035
- Payments	17,544	20.1%	1,830M	104,309
- Insurance	4,999	5.7%	397M	79,416
- Investment	4,325	4.9%	251M	58,035
- Credit, excl. personal loans	16,252	18.6%	1,290M	79,375
- Personal loans	5,106	5.8%	826M	161,770

Suggestive evidence of problems in personal lending apps (1/2)

We observe high "churn rates" (i.e., new entry and exits) among personal lending apps relative to other finance app categories.

Table: Percentage of finance apps newly-released & unpublished during study period (Jan 2020-April 2021)

	App product category						
	(1) **All**	(2) Banking	(3) Payments	(4) Insurance	(5) Investment	(6) General credit	(7) **Personal loan**
Panel A. Entry and exit							
% New release	**34%**	22%	43%	30%	35%	39%	**70%**
% Unpublished	**24%**	13%	25%	20%	20%	26%	**52%**
Panel B. Time on app store							
Avg. app age (in mths.)	**29.3**	41.0	24.8	27.9	27.5	27.5	**11.4**
Observations	**134,744**	18,015	17,544	4,999	4,325	16,252	**5,106**

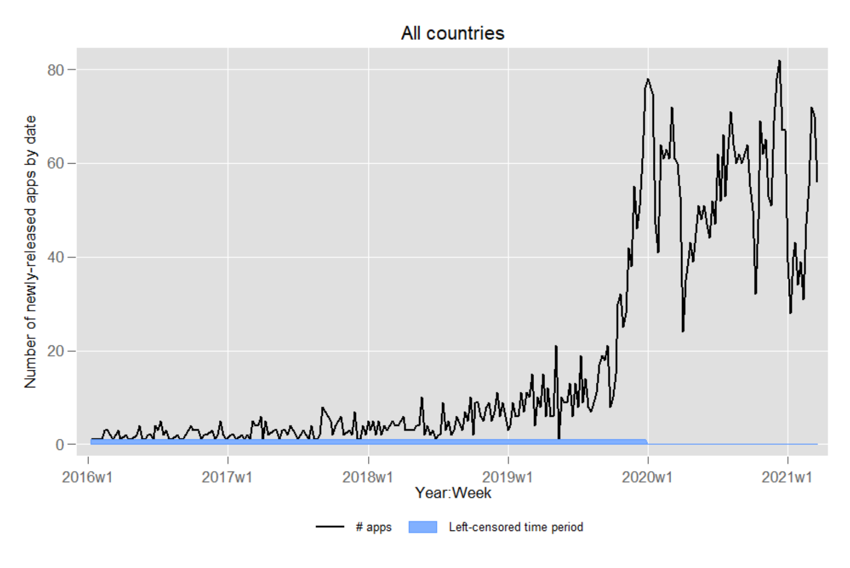
Suggestive evidence of problems in personal loan apps (2/2)

- "Churn" fairly consistent across different country markets. Signs many apps operate in multiple markets.
- Removals increase another ≈10% as we tag up until Sep. 2021

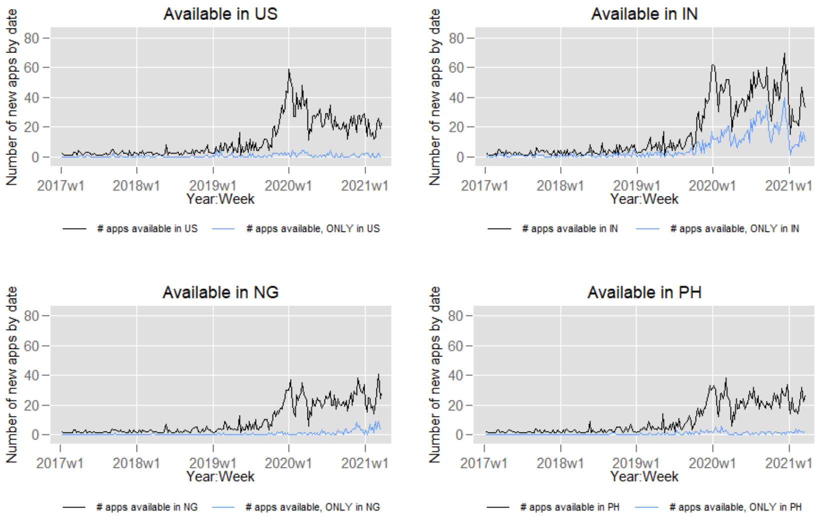
Table: Percentage of personal loan apps newly-released & unpublished during extended period (Jan 2020-Sep 2021)

		App market availability			
	(1) Personal loan apps	(2) Available in US	(3) Available in IN	(4) Available in NG	(5) Available in PH
New release	71%	65%	71%	69%	67%
Unpublished by Apr. 2021	52%	54%	56%	46%	47%
Unpublished by Sep. 2021	63%	65%	67%	61%	61%
Observations	5,106	2,613	3,810	2,281	2,241

Proliferation of personal lending apps in aggregate: supply side



Proliferation of personal lending apps by country: supply side



A typology of personal lending apps

We systematically review the 5,106 pure-play personal lending (or related) apps. We identify several broad types (and subtypes) of lending apps:

A. DIRECT PERSONAL LENDING APPS (74.9%)

- Pure-play lending apps that claim to directly provide personal loans (including payday, consumer, MSME, etc.)
- Any provider type (e.g., commercial bank, neobank, fintech, MFI, NBFI, others)

B. INDIRECT PERSONAL LENDING APPS (19.8%)

- Indirect lender apps that do not directly offer loans, but serve as “marketplace/guide” to access personal loans
- In practice, some are more or less transparent about their intermediary role

C. PERIPHERAL APPS RELATED TO PERSONAL LENDING (5.3%)

- Loan / EMI calculators
- Credit report / score apps

1. “Pure fraud” app example (from India)

Characterized by existence of fake reviews.

Rupee

India-credit Provider Finance

★★★★★ 927

PEGI 3

You don't have any devices

Add to Wishlist

Rupee is a mobile financial platform, which can provide users with the safe loan service. The new mode of simple, convenient, flexible loan eliminates the tedious process and application program.

Eligibility
1. Indian citizen
2. 18-55 years old
3. Have a stable income

Features:
Loan Amount: from ₹2000 to 200000
Tenure: Minimum of 91 days and Maximum of 365 days
Maximum Annual Percentage Rate(APR): 0.08% per day, also is 29.2% per annum.
Processing Fee: Up to 5% of Loan Amount (additional 18% GST)

Example:
For ₹10000 loan of 91 days duration
Applicable fees: ₹10000 * 5% = ₹500
GST: ₹500 * 18% = ₹90
Interest payable: ₹10000 * 29.2% / 365days * 91days = ₹728
Total cost of the loan: ₹500 + ₹90 + ₹728 = ₹1318
Total repayment amount(91days): ₹10000 + ₹1318 = ₹11318

WHAT'S NEW
Fix bugs

ADDITIONAL INFORMATION

Updated	Size	Installs
September 14, 2021	5.1M	50,000+
Current Version	**Requires Android**	**Content Rating**
1.0.15	5.0 and up	PEGI 3 Learn more
Permissions	**Report**	**Offered By**
View details	Flag as inappropriate	India

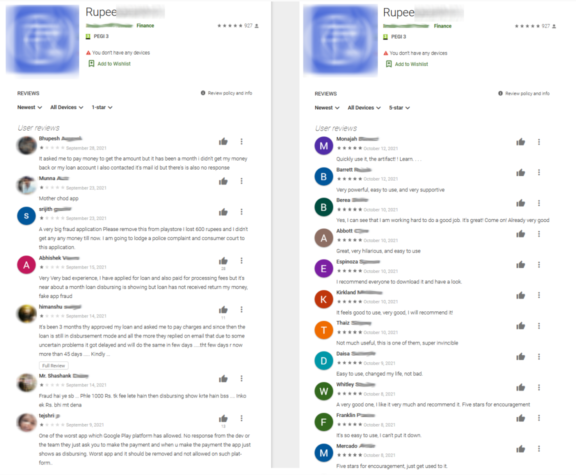
Developer
indu@gmail.com
Privacy Policy

▶▶ To example of scam website

Classification of suspect apps (1/3)

1. “Pure fraud” app example (from India)

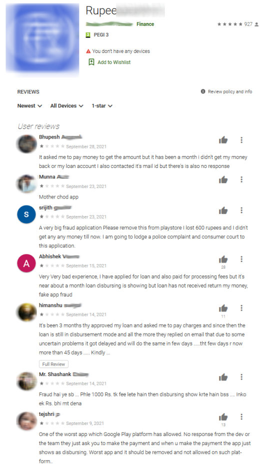
Characterized by existence of fake reviews.

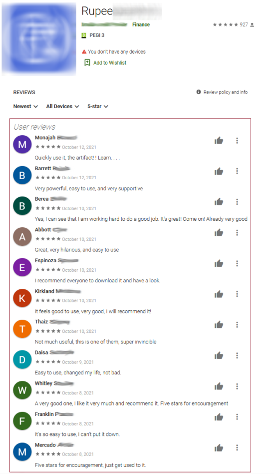


Classification of suspect apps (1/3)

1. “Pure fraud” app example (from India)

Characterized by existence of fake reviews.





Examples of signals

- Idiosyncratic review dates
- Abnormal user names
- Idiosyncratic text

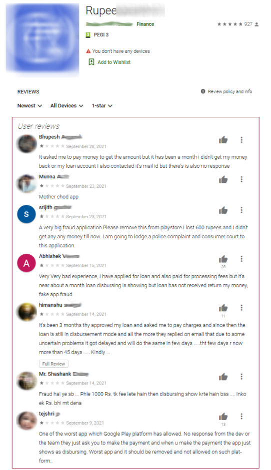
Classification of suspect apps (1/3)

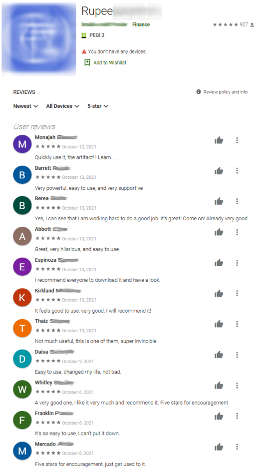
1. “Pure fraud” app example (from India)

Characterized by existence of fake reviews.

Examples of signals

- Regularly distributed review dates
- Complaints about app/reviews being fake
- Paid “registration fees” but no service





Signs of systematized (fake) reviewers / review farms

Table: Reviews across multiple personal lending apps for suspected (fake) reviewer

package_name	*reviewer_name*	*profile_pic*	*rating*	*review_date*	*review_body_text*
com.pearloans.#####	ajayi #####	0	5	23-Jun-21	Very reliable and trustworthy
com.smartloans.#####	ajayi #####	0	5	28-Jun-21	Very reliable and trustworthy
com.flycashloans.#####	ajayi #####	0	5	1-Jul-21	Very reliable and trustworthy
com.loan#####	ajayi #####	0	5	1-Jul-21	Very reliable and dependable
com.liberty#####	ajayi #####	0	5	1-Jul-21	Very reliable and trustworthy
com.oc.ourr_#####	ajayi #####	0	5	1-Jul-21	Very reliable and dependable
com.ttitocash.#####	ajayi #####	0	5	3-Jul-21	Very reliable and trustworthy
com.value#####	ajayi #####	0	5	4-Jul-21	Very reliable and dependable

Classification of suspect apps (2/3)

2. “Predatory” app example (from Nigeria)

Characterized by signs of i) no real services provided or ii) abusive lending practices.

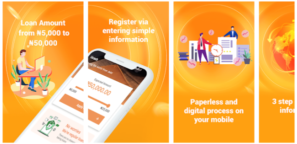
-Personal Loan Online , 7*24 hour

Loan Finance ★★★★★ 469

PEGI 3

You don't have any devices

Add to Wishlist



makes it efficient for you to access credit anytime, anywhere.

- ✔ Loans Online amount: from ₦5,000 to ₦50,000
- ✔ Loans period range from 91 to 180 days
- ✔ Annual interest rates from 36.5% to 300%. The interest rate is calculated on a daily basis.
- ✔ Flexible repayment Plan
- ✔ Interest and service rate: With a minimum of 0.1% up to 1%

→ Example:
Borrow NGN50,000 over 3 months
Interest (total cost of the loan): NGN19,000 (NGN50,000 * 38% = NGN19,000 interest)
Representative APR: 120%

Total amount to repay will be NGN 69,000
Three monthly repayments of NGN 23,000

→ Loan Applicant Minimum Requirement
✓ Nigeria Resident
✓ Age above 20 and below 55

→ How to get Started.
To apply for online loans on app you need an android phone and Internet.
We are working hard to provide you with the best services to make your life better. Very soon you will look to for your electricity bill payments, current account, savings and much more!

WHAT'S NEW
Thank you users, for your support.
We have launched the Golden Egg Smashing event, welcome to participate!

ADDITIONAL INFORMATION

Updated	Size	Installs
April 26, 2021	9.1M	50,000+
Current Version	**Requires Android**	**Content Rating**
1.4	4.4 and up	PEGI 3 Learn more
Permissions	**Report**	**Offered By**
View details	Flag as inappropriate	Loan

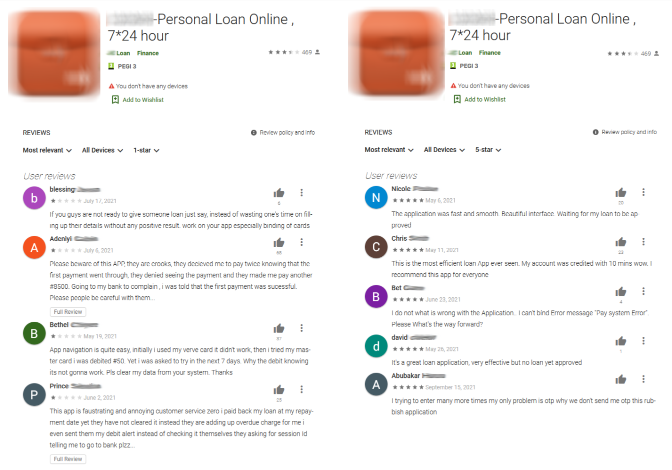
Developer
ic79@gmail.com
Privacy Policy

⏩ To example of scam website

Classification of suspect apps (2/3)

2. “Predatory” app example (from Nigeria)

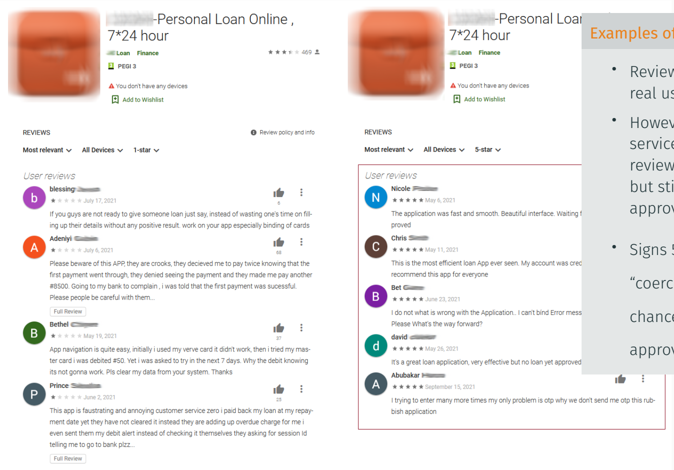
Characterized by signs of i) no real services provided or ii) abusive lending practices.



Classification of suspect apps (2/3)

2. “Predatory” app example (from Nigeria)

Characterized by signs of i) no real services provided or ii) abusive lending practices.



Examples of signals

- Reviews seem from real users
- However, no signs services provided (e.g., review seems positive but still awaiting approval)
- Signs 5-star ratings “coerced” to increase chances of loan approval

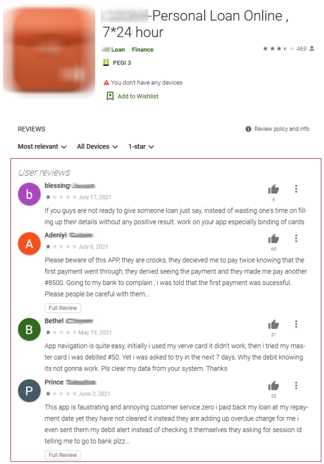
Classification of suspect apps (2/3)

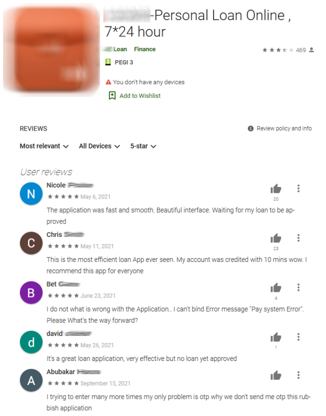
2. “Predatory” app example (from Nigeria)

Characterized by signs of i) no real services provided or ii) abusive lending practices.

Examples of signals

- Consistent signs services not provided
- Applications submitted (often including PII) but no approvals
- Sometimes complaints over fees or unsanctioned debits





Classification of suspect apps (3/3)

3. “Predatory” app example (from the Philippines)

Characterized by signs of i) no real services provided or ii) abusive lending practices.

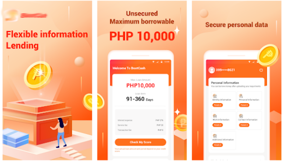
Cash

Eastman Finance ★★½ 1,072

PEGI 3

You don't have any devices

Add to Wishlist



develops new loan products
• Maximum amount ₱10000
• Loan Term: 91days-365days
• APR: 10.8%
• Service fee: 0.2%
• Transaction fee: 0
Example: ₱10000 for 91 days (interest 0.03%daily)
₱10000/91=10000*（1+0.03%）^91+10000*0.2%=₱10296
Total 91days ₱10296

An online lending platform for loans in the Philippines that provides Filipinos with secure and affordable fast cash online lending services for peso loans.

Easy to apply: Fill out an application online in just a few steps without having to sign a paperless contract
Payment: After approval, just wait 1 hour to 2 hours to pay

Low interest rate: For those who have a good track record on , they will get the lowest interest rate.
Faster refinancing: Once the customer pays off the loan, you will be eligible to refinance
Quick Help: A customer service representative will resolve all your loan issues immediately
Safe and reliable: will protect and keep all customer information confidential

WHAT'S NEW
Update to new version

ADDITIONAL INFORMATION

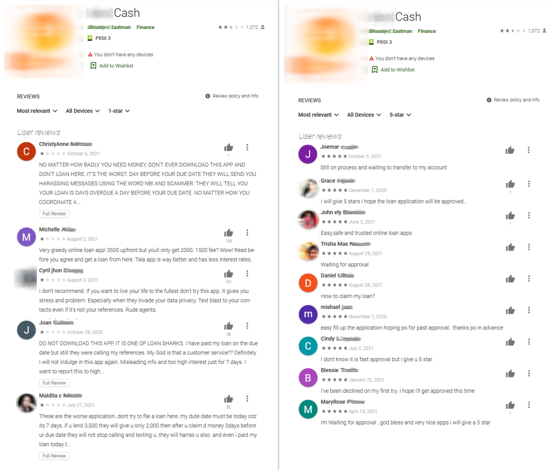
Updated	Size	Installs
August 20, 2021	7.0M	500,000+
Current Version	**Requires Android**	**Content Rating**
1.0.7	5.0 and up	PEGI 3 Learn more
Permissions	**Report**	**Offered By**
View details	Flag as inappropriate	Eastman

Developer
manb71@gmail.com
Privacy Policy

Classification of suspect apps (3/3)

3. “Predatory” app example (from the Philippines)

Characterized by signs of i) no real services provided or ii) abusive lending practices.



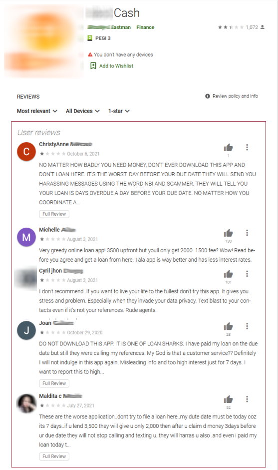
Classification of suspect apps (3/3)

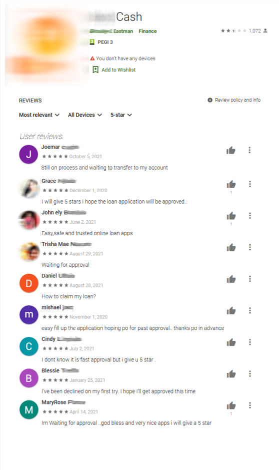
3. “Predatory” app example (from the Philippines)

Characterized by signs of i) no real services provided or ii) abusive lending practices.

Examples of signals

- High prevalence of complaints on interest rates, fees, or short tenure
- Common complaints on abusive debt collection practices
- Concerns over data privacy

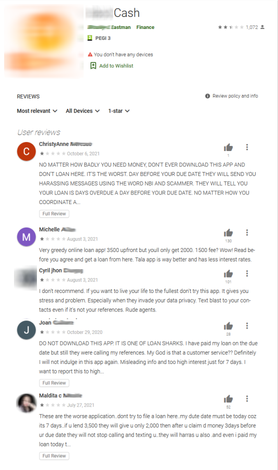


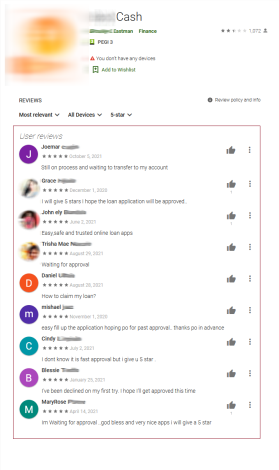


Classification of suspect apps (3/3)

3. “Predatory” app example (from the Philippines)

Characterized by signs of i) no real services provided or ii) abusive practices.

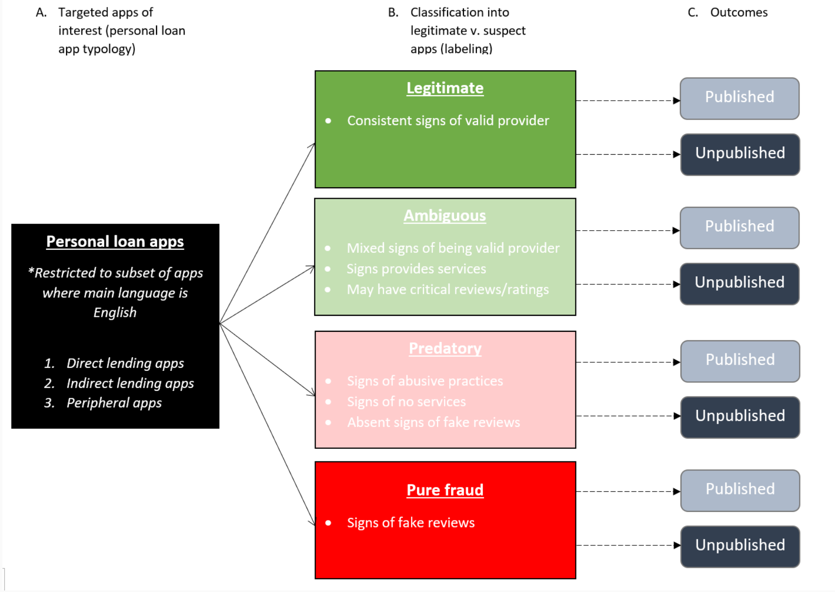




Examples of signals

- Sometimes signs of “coerced” 5-star ratings, where mismatch between rating and tone of review
- Signs that some form of services provided for some users

Recap: overview of typology and classification



Outcomes for legitimate vs. suspect personal lending apps

	Prevalence to be "unpublished" (as of Apr. 2021)
# of all personal loan apps	5, 106
4-category classification	
% Legitimate	16.4%
% Ambiguous	9.3%
% Predatory	70.1%
% Pure fraud	7.7%
Binary classification	
% Likely legitimate	13.8%
% Likely suspect	67.2%

Estimates on prevalence of legitimate vs. suspect personal lending apps (1/3)

	Prevalence of app in category				
	All (1)	Available in US (2)	Available in IN (3)	Available in NG (4)	Available in PH (5)
Panel A. Number of personal lending apps					
# of apps	5,106	2,613	3,810	2,281	2,241
Panel B. Percent of apps falling in legitimate vs. suspect classification buckets					
% in 4-category classification					
- **Legitimate**	16.8%	16.6%	15.7%	16.4%	16.9%
- **Ambiguous**	9.9%	11.1%	9.2%	11.7%	12.2%
- Predatory	68.8%	69.1%	71.4%	68.1%	67.2%
- Pure fraud	4.6%	3.3%	3.7%	3.9%	3.8%
% in Binary classification					
- **Likely legitimate**	26.7%	27.7%	25.0%	28.1%	29.0%
- Likely suspect	73.3%	72.3%	75.0%	71.9%	70.9%

Estimates on uptake of legitimate vs. suspect personal lending apps (2/3)

	Estimated downloads in millions (M)				
	All (1)	Available in US (2)	Available in IN (3)	Available in NG (4)	Available in PH (5)
All personal lending apps	834M	418M	670M	394M	407M
4-category classification					
- **Legitimate**	616M	369M	522M	343M	344M
- **Ambiguous**	38M	17M	29M	16M	27M
- Predatory	149M	24M	106M	26M	29M
- Pure fraud	31M	8M	14M	9M	8M
Binary classification					
- **Likely legitimate**	654M	386M	551M	359M	371M
- Likely suspect	180M	32M	119M	35M	36M

**Note: these values reflect cumulative downloads since initial release of apps. For some apps, this may include downloads prior to Jan. 1st, 2020.*

Estimates on average uptake of legitimate vs. suspect personal lending apps (3/3)

	Average downloads per app in category				
	All (1)	Available in US (2)	Available in IN (3)	Available in NG (4)	Available in PH (5)
All personal lending apps	163,337	159,969	175,853	172,731	181,615
4-category classification					
- **Legitimate**	716,279	850,230	871,452	914,667	910,053
- **Ambiguous**	76,342	59,170	82,102	59,023	97,802
- Predatory	42,246	13,172	38,757	16,582	18,857
- Pure fraud	144,651	116,020	109,756	124,801	113,034
Binary classification					
- **Likely legitimate**	479,824	533,887	579,390	560,062	569,892
- Likely suspect	48,103	16,781	41,638	21,293	22,782

**Note: these values reflect cumulative downloads since initial release of apps. For some apps, this may include downloads prior to Jan. 1st, 2020.*

Adverse impacts on consumers

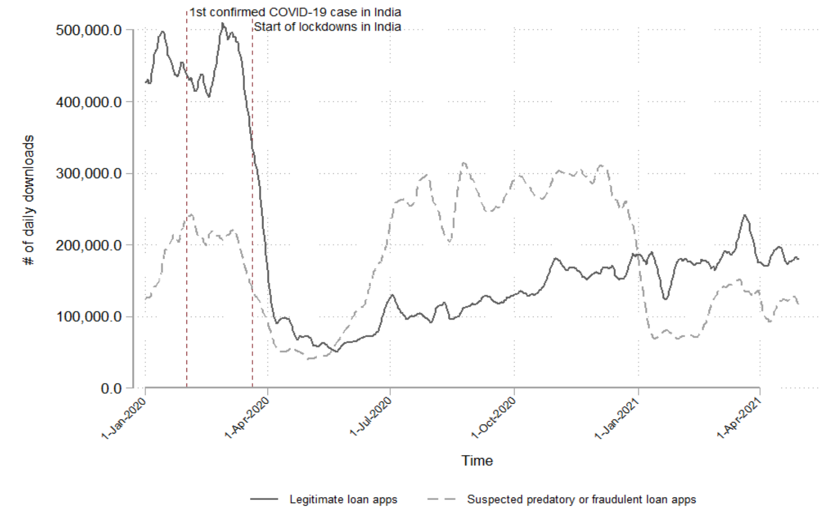
Measurable harm

- Financial loss due to direct fraud (e.g., payment of fees for non-existent services)
- Financial loss due to predatory practices (e.g., difference in fees and interest paid between predatory lender relative to similar legitimate lenders serving market)

Other adverse outcomes / Immeasurable

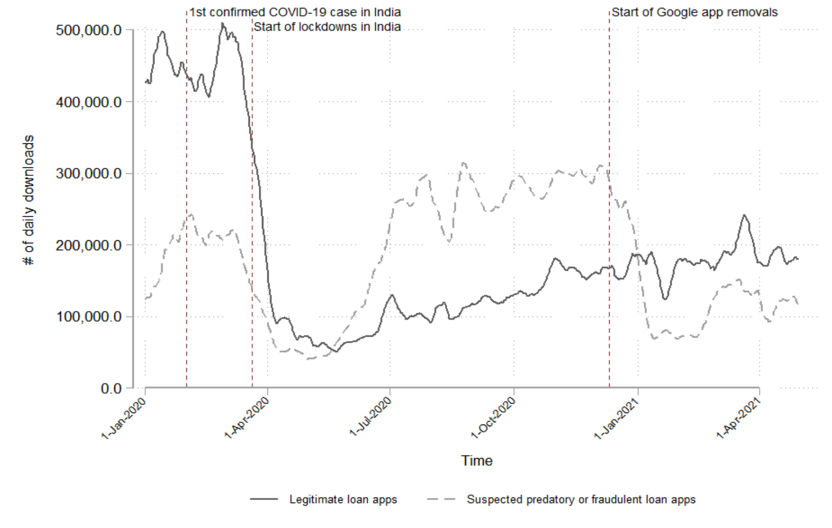
- Loss of private / sensitive data
- Harassment or abuse encountered
- Loss of trust in financial institutions

Negative externalities on legitimate providers?



▸▸ Examples from other markets

Policy solutions have had shorter-term success (but insufficient..)



Examples from other markets

The proposed solution

Solution summary

1. We set up ML models as proof-of-concept of proposed solution

2. The labeled data (for the 5,106 personal loan apps) serve as main output variables. We run the following models:
    - **Binary classification model:** i) “likely legitimate” vs. ii) “likely suspect”
    - **3-category classification model:** i) “likely legitimate” vs. ii) “predatory” vs. iii) “pure fraud”

3. For model input variables, we prioritize use of app meta-data
    - Motivation is to see if these mostly static and ex-ante indicators can flag problematic apps
    - We intend to integrate review data, but these data can be noisy and imply being ex-post after problems have arisen

4. In practice, we split the sample into a training-validation and test datasets
    - **Training-validation dataset** uses apps released until Jan. 15th, 2021
    - **Test dataset** uses apps released after Jan. 15th, 2021 and is used to assess out-of-sample performance of the model on completely unseen data

Input variables summary

Based on our exploration, we use a subset of meta data variables in our current models. These include both visible and hidden data and often imply various transformations.

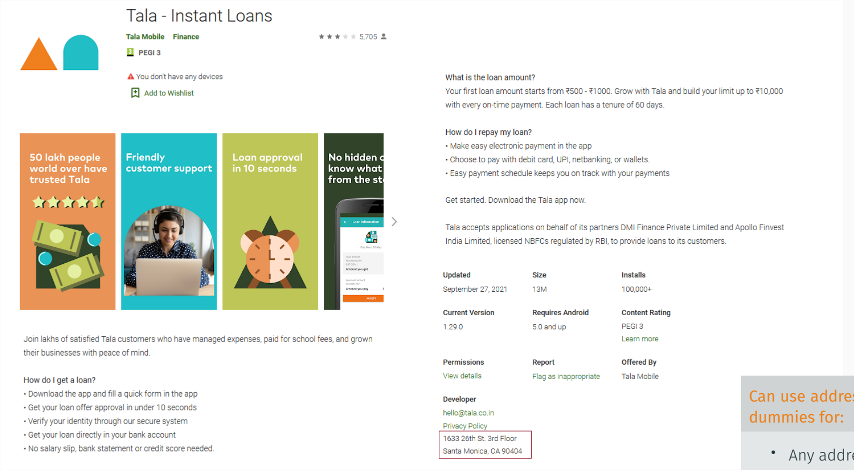
Visible data

Short/long text description
Screenshots
Date last updated
Byte size
Developer website
Developer email
Promo video

Hidden data

Date created
Disaggregated # ratings (1-5)
Internal version code
Whether contains ads
Number of reviews
App price
In-app purchases
Min/Max price for in-app purchases
Interactive elements
Countries available in
Languages available in
Other app-stores available in
Downloads (country-day level)

Input variable transformation: Example 1



Can use address variable to construct dummies for:

- Any address? (0/1)

Input variable transformation: Example 2

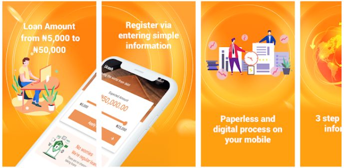
-Personal Loan Online , 7*24 hour

Loan Finance ★★★½ 469

PEGI 3

You don't have any devices

Add to Wishlist



makes it efficient for you to access credit anytime, anywhere.

- ✔ Loans Online amount: from ₦5,000 to ₦50,000
- ✔ Loans period range from 91 to 180 days
- ✔ Annual interest rates from 36.5% to 300%. The interest rate is calculated on a daily basis.
- ✔ Flexible repayment Plan
- ✔ Interest and service rate: With a minimum of 0.1% up to 1%

→ Example:
Borrow NGN50,000 over 3 months
Interest (total cost of the loan): NGN19,000 (NGN50,000 * 38% = NGN19,000 interest)
Representative APR: 120%

Total amount to repay will be NGN 69,000
Three monthly repayments of NGN 23,000

→ Loan Applicant Minimum Requirement
✓ Nigeria Resident
✓ Age above 20 and below 55

→ How to get Started.
To apply for online loans on app you need an android phone and Internet.
We are working hard to provide you with the best services to make your life better. Very soon yo
look to for your electricity bill payments, current account, savings and much more!

WHAT'S NEW
Thank you users, for your support.
We have launched the Golden Egg Smashing event, welcome to participate!

ADDITIONAL INFORMATION

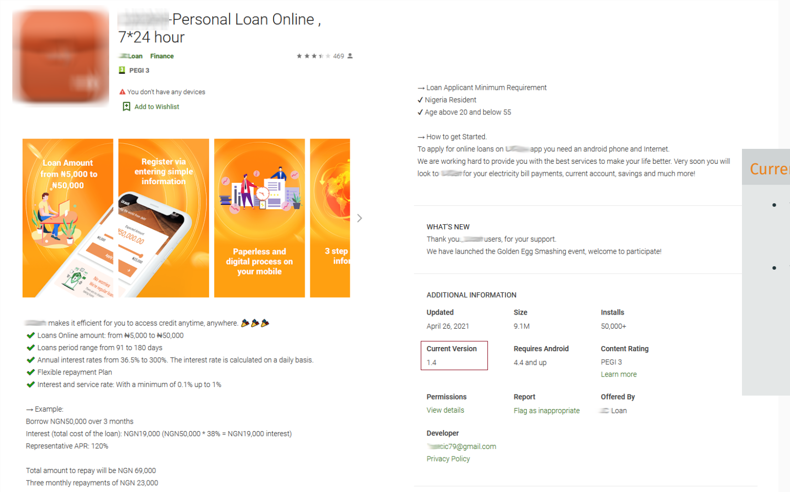
Updated	Size	Installs
April 26, 2021	9.1M	50,000+
Current Version	**Requires Android**	**Content Rating**
1.4	4.4 and up	PEGI 3 Learn more
Permissions	**Report**	**Offered By**
View details	Flag as inappropriate	Loan

Developer
ic79@gmail.com
Privacy Policy

Can use email address variable to construct dummies:

- Any email provided? (0/1)

Input variable transformation: Example 3



Current version vs. version code:

- The **current version** listed on the url is set by the developer.
- The meta-data has a field **version code**, which reflects the number of iterations of the app. We use this directly.

Methodology

1. We use an ML technique called "gradient boosting" (specifically *xgboost*)

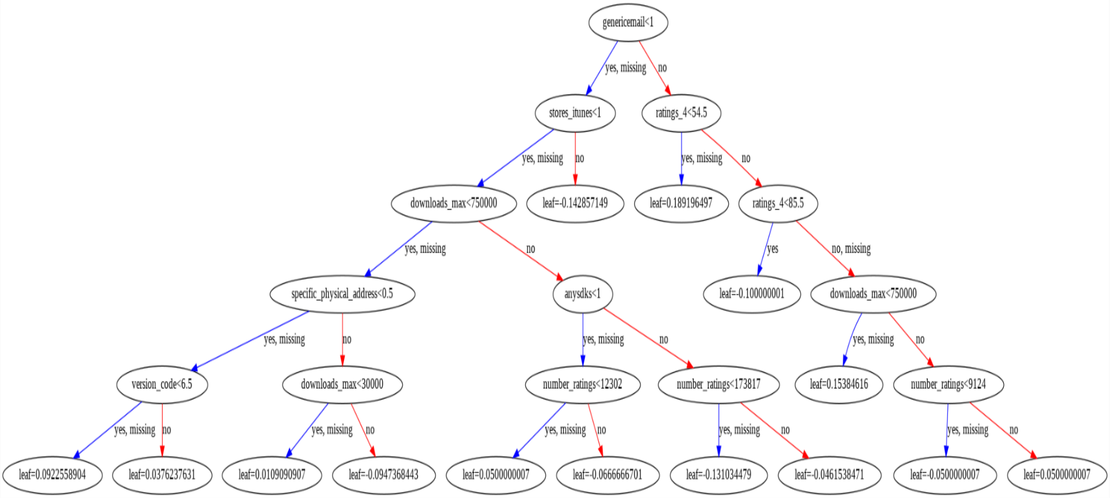
2. *XGboost* considered state-of-art for many predictive classification tasks
   - It builds an ensemble of shallow and weak successive trees with each tree learning and improving on the previous
   - When combined, these many weak successive trees produce a powerful "committee" that are often hard to beat with other algorithms

3. Key advantages:
   - Generally more accurate compare to other modes
   - Lots of flexibility – can optimize on different loss functions and provides several hyperparameter tuning options that make function fit very flexible
   - Handles missing data - imputation not required

4. Some disadvantages:
   - More likely to overfit
   - Training time scales with data size

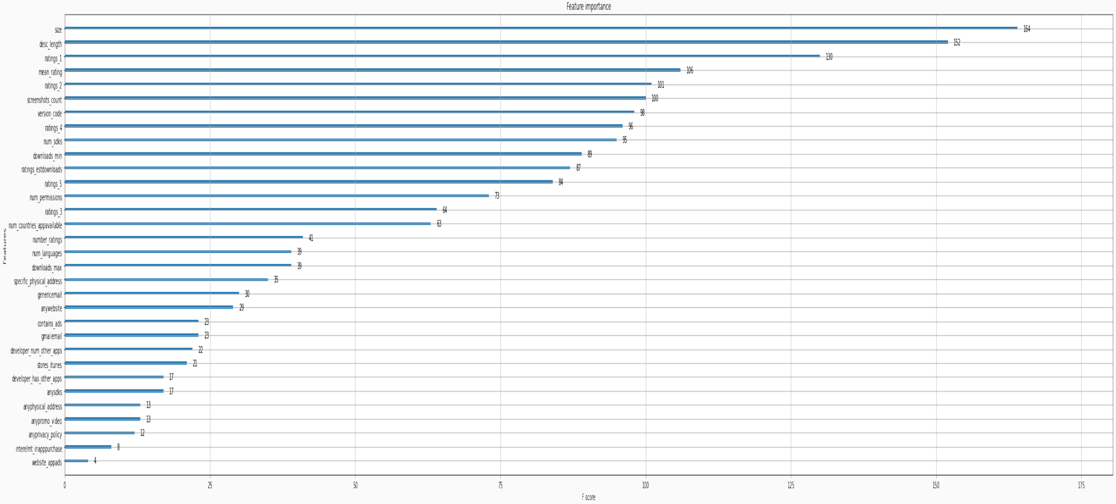
Results for binary classification model (1/2)

Figure: Decision tree



Results for binary classification model (2/2)

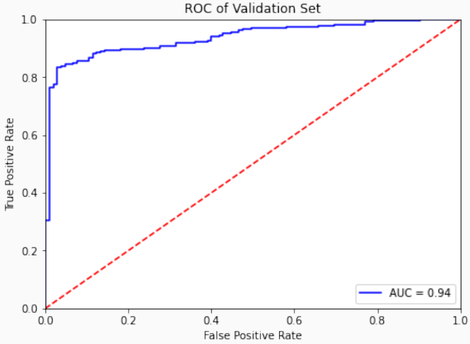
Figure: Feature importance



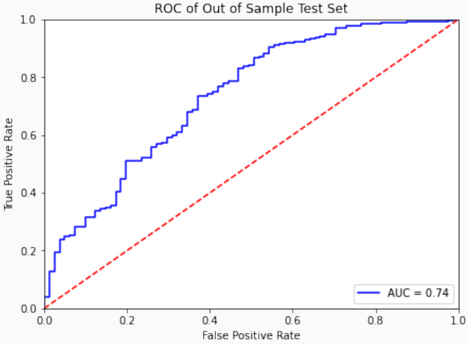
Accuracy for binary classification

	Validation Set	Out of Sample
% Accuracy	90.17%	84.45%

Evaluation of binary classification model performance – ROC curve

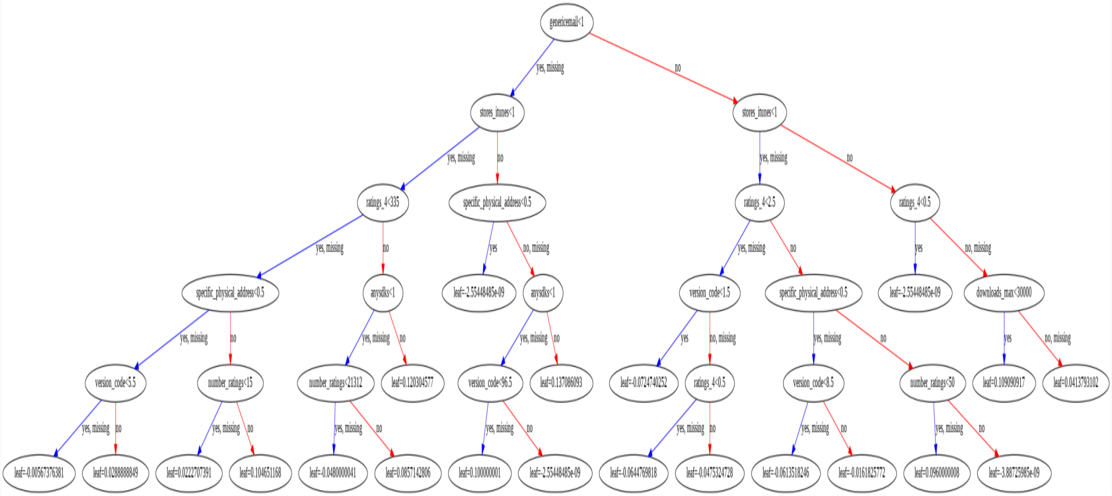


ROC for validation sample



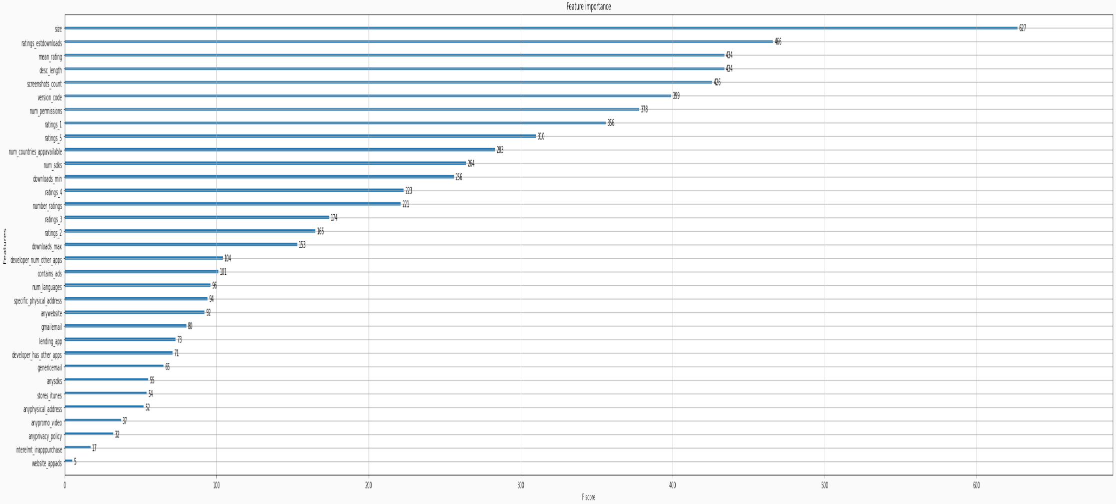
ROC for test sample

Figure: Decision tree



Results for binary classification model (2/2)

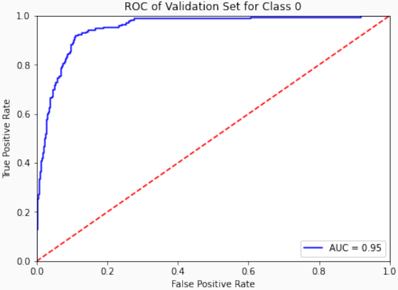
Figure: Feature importance

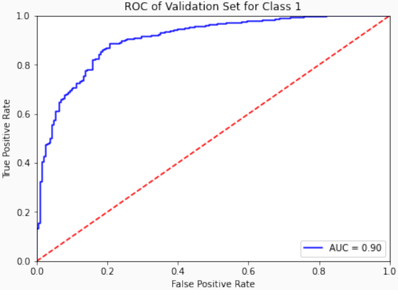


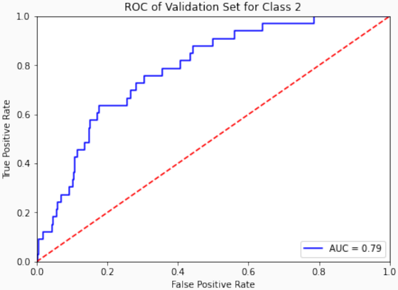
Accuracy for multi-class classification

	Validation Set	Out of Sample
% Accuracy	85.61%	71.96%

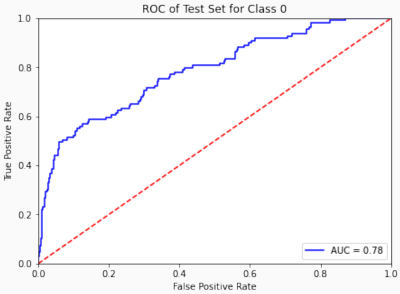
Evaluation of multi-class classification model performance – ROC curves (1/2)

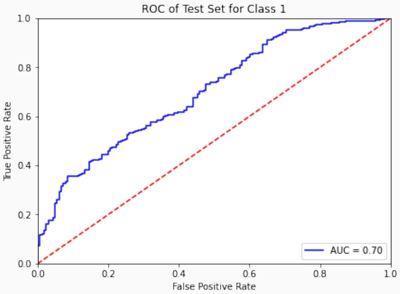


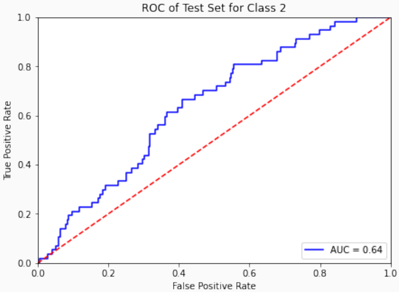




Evaluation of multi-class classification model performance – ROC curves (1/2)





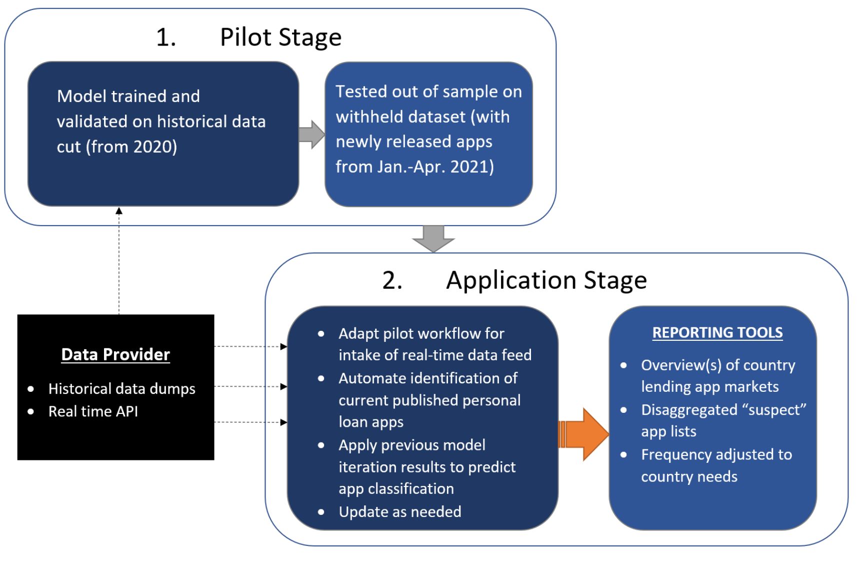


Main takeaways & options for improving performance

1. Increase data sample size (e.g., extend prior to Jan 2020 or after April 2021)
2. Model-specific adjustments: e.g., further variable construction/transformation, hyperparameter tuning, etc.
3. Test whether narrowing windows between train-validation and out-of-sample increases precision (i.e., by potentially comparing more similar apps)
4. Test how much integration of (ex-post) review data improves performance

Next Steps & Conclusion

Moving from a proof-of-concept to real-time tool



Next steps

1. Integrate review data into analysis and ML models
2. Further validate results by downloading and using random subset of apps
3. Systematize evidence on adverse consumer outcomes
4. Expand coverage into other finance app categories and non-English country markets

Q & A / Feedback

Feedback

We'd like to obtain feedback from possible end users on how these findings and analyses may fit or be better tailored to applied policy, regulatory, or industry needs – e.g.:

1. Is interest more in ex-ante or ex-post tools?
    - If former, what is acceptable tolerance for error margin? Or leverage in requiring more stringent vetting process?
    - If latter, what type of reporting tools and info most desired?
2. Frequency of desired reporting?
3. Expand into other country markets?
4. Expand into other non-lending product coverage?
5. Other comments or suggestions?