

# اعتبارسنجی در لندتک با استفاده از داده کاوی

سیدحسین سیداصفهانی  
رئیس هیات مدیره لندو

# هدف

انتخاب ویژگی و کشف الگو، به منظور افزایش دقت اعتبارسنجی مشتریان لندتک

# اعتبارسنجی

- اعتبار سنجی پایه و اساس تصمیم گیری اعطای اعتبار در لندتک می باشد.
- یک مدل اعتبارسنجی کارا و مناسب می تواند باعث جلوگیری از صرف هزینه های گزاف برای لندتک و کاهش ریسک اعتباری تامین کننده مالی شود.
- اعتبارسنجی، مجموعه ای از مدل های تصمیم گیری و تکنیک های اساسی آن می باشد که به وام دهندگان، جهت اتخاذ تصمیمات مناسب در اعطای اعتبار یاری می رساند.

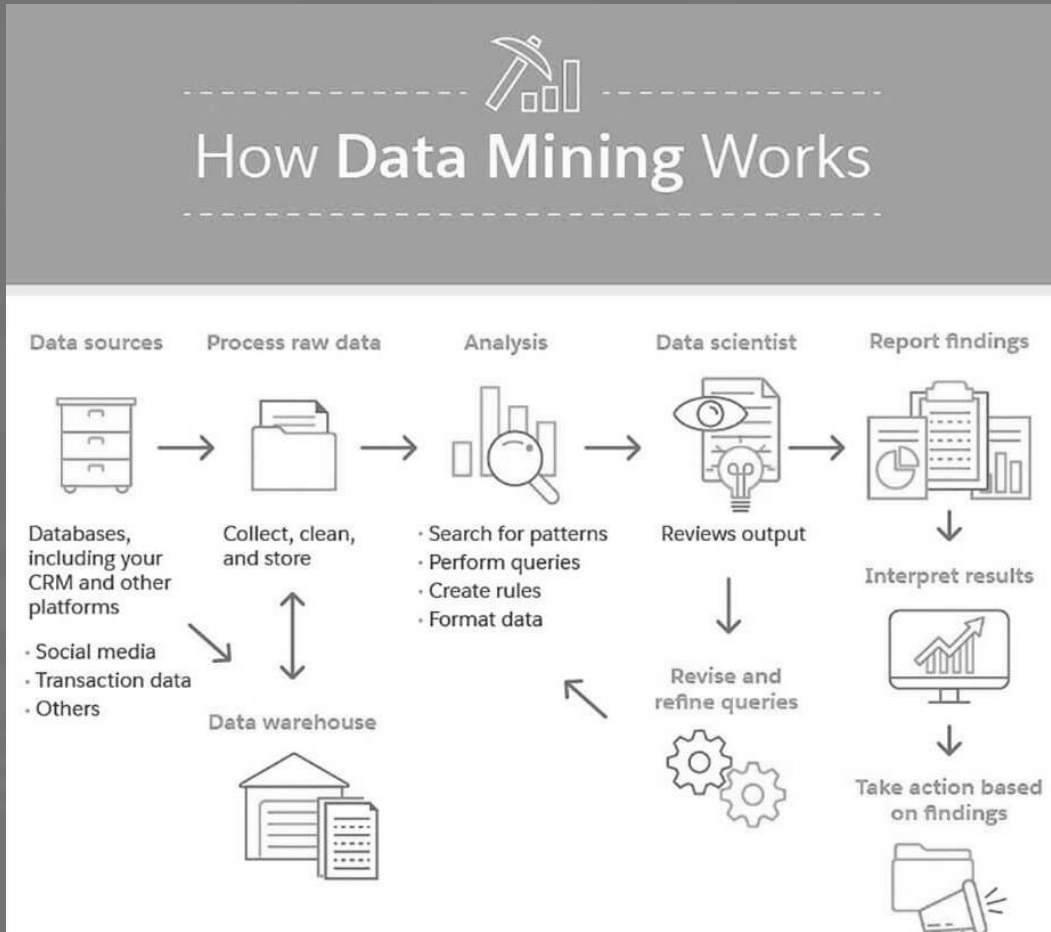
# مشتریان خاص لندتک

- حدود ۷۰٪ از مشتریان لندتک جوانان با میانگین سنی ۱۸-۲۳ سال هستند که به اینترنت دسترسی داشته و سواد استفاده از پلتفرم های دیجیتال را دارند.
- وام اولی ها و بدون سابقه رفتار بازپرداخت تسهیلات.
- مشتری سایر پلتفرم ها و فعال در شبکه های اجتماعی.
- ...

# داده کاوی

داده کاوی عبارت است از فرآیند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده های بزرگ، و استفاده از آن در تصمیم گیری

یادگیری اطلاعات  
در مورد رفتار مشتری



# انتخاب ویژگی‌ها

- فرایند انتخاب ویژگی‌ها، از مسائل بهینه سازی ترکیبی کلی در یادگیری ماشین شناخته می‌شود که تعداد ویژگی‌ها را کاهش داده و داده های غیر مرتبط و زائد را حذف می‌کند.
- هدف اصلی انتخاب ویژگی، شناسایی زیرمجموعه‌ای از ویژگی‌های بهینه است که بیشترین تأثیر را دارند.



*Research article*

## Machine learning and artificial neural networks to construct P2P lending credit-scoring model: A case using Lending Club data

An-Hsing Chang<sup>1</sup>, Li-Kai Yang<sup>2</sup>, Rua-Huan Tsaih<sup>3</sup> and Shih-Kuei Lin<sup>1,\*</sup>

<sup>1</sup> Department of Money and Banking, National Chengchi University, Taiwan

<sup>2</sup> Department of Investment and Trading, Cathay Life Insurance Co., Ltd., Taiwan

<sup>3</sup> Department of Management Information Systems, National Chengchi University, Taiwan

\* Correspondence: Email: [square@nccu.edu.tw](mailto:square@nccu.edu.tw)

**Abstract:** In this study, we constructed the credit-scoring model of P2P loans by using several machine learning and artificial neural network (ANN) methods, including logistic regression (LR), a support vector machine, a decision tree, random forest, XGBoost, LightGBM and 2-layer neural networks. This study explores several hyperparameter settings for each method by performing a grid search and cross-validation to get the most suitable credit-scoring model in terms of training time and test performance. In this study, we get and clean the open P2P loan data from Lending Club with feature engineering concepts. In order to find significant default factors, we used an XGBoost method to pre-train all data and get the feature importance. The 16 selected features can provide economic implications for research about default prediction in P2P loans. Besides, the empirical result shows that gradient-boosting decision tree methods, including XGBoost and LightGBM, outperform ANN and LR methods, which are commonly used for traditional credit scoring. Among all of the methods, XGBoost performed the best.

**Keywords:** P2P lending; credit score; machine learning; artificial neural networks; feature engineering; Lending Club

Version of Record: <https://www.sciencedirect.com/science/article/pii/S0167268120300822>  
Manuscript\_b2f7d0fa1411298a49ad20ca353b4051

## Important Factors Determining Fintech Loan Default: Evidence from a LendingClub Consumer Platform

Christophe Croux<sup>1</sup>  
EDHEC Business School  
Julapa Jagtiani<sup>2</sup>  
Federal Reserve Bank of Philadelphia  
Tarunsai Korivi<sup>3</sup>  
Amazon.com  
Milos Vulcanovic<sup>4\*</sup>  
EDHEC Business School

January 27<sup>th</sup>, 2020

### Abstract

This study examines the default determinants of Fintech loans, utilizing a sample of more than a million of personal loans that were originated through the LendingClub consumer platform during the period 2007-2018. We identify a robust set of contractual loan characteristics, borrower characteristics, and macroeconomic variables that are important in determining the likelihood of default, such as loan maturity, homeownership, loan purposes, occupation, etc. We also find an important role of alternative data in determining the default, even after controlling for the obvious risk characteristics of the borrowers, loan characteristics, and the local economic factors. The results are robust to different empirical approaches. Results imply that it would be important for regulators to provide greater transparency in terms of guidance and regulatory clarity on which alternative data can be used legally without violating fair lending rules. Lenders need to pay closer attention to how they make decisions and understand their own decisions that may be driven by complex algorithms inside the “black boxes.”

**Keywords:** Big data, Crowdfunding, Financial innovation, Household Finance, Lasso selection methods, Machine learning, Peer to peer lending, P2P/marketplace lending