BIG DATA
SCORING

Digital Footprint

In-depth product overview

CONTENTS

PRODUCT OVERVIEW

Digital Footprint LITE and PREMIUM are the advanced tools for lenders to improve their credit quality and loan acceptance rates through use of big data. The solutions help lenders reduce 3 of the most common mistakes when issuing loans:

false positives

lending to not creditworthy applicants

false negatives

declining loan to creditworthy applicants

fraud

lending to people who have no intention of repayment

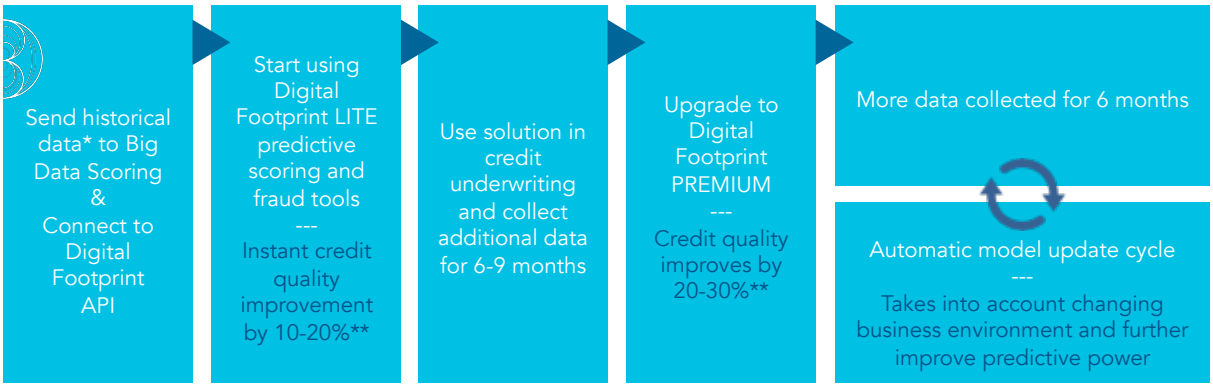
The solutions collect up to 5,000 additional data points for potential borrowers from a variety of sources that are not used by most lenders today. The new data points are then processed by our big data scoring algorithms to exactly predict each loan applicant's probability of default (the Digital Footprint Credit Score). When combined with the in-house underwriting processes, the Digital Footprint Credit Scores improve the scoring accuracy by up to 30%, which directly translates into more revenue and better credit quality for lenders (see case studies on next pages).

The LITE version of the solution is immediately available for all lenders that sign up with Big Data Scoring. Lenders whose client acquisition channels are mostly online have the possibility to upgrade to the PREMIUM solution in 6-9 months and hence improve scoring accuracy even further. Both versions of the solution also include advanced Fraud Detection tools that help verify the identity of the loan applicant and detect suspicious activity on the lender's web site.

Digital Footprint is constantly updated to capture more data, find new ways to understand the data, recognize fraudulent behaviour and adapt to emerging fraud patterns.

ROADMAP

In order to start using the LITE solution, a lender has to send a historical dataset of its clients to Big Data Scoring. The dataset is used for Digital Footprint model calibration for the lender's needs and business specifics. As a result, the lender will get access to a bespoke machine-learning algorithm (Digital Footprint LITE) that significantly enhances the predictive power of its in-house underwriting process.



** - minimum 10,000 sample with certain input information (consult with BDS for details)*
*** - exact improvement figures depend on the quality of in-house underwriting models, local credit bureau information, credit product, marketing channels and other factors*

After using the LITE solution for 6-9 months together with additional data collection tools provided by Big Data Scoring, we can upgrade the algorithms to use even more data in the calculation of expected payment behavior and the lender would get access to the PREMIUM solution.

USING THE SOLUTIONS

After development of the bespoke credit prediction algorithms, a scorecard is presented to the lender showing the predictive power of the Digital Footprint solution. The scorecard divides all loan applicants into a number of scorebands (usually 8-10) and highlights the expected default probability of each scoreband (e.g. scoreband number 1 might have an expected default probability of 74.3% and the respective figure for scoreband 8 might be 24.1%). The Digital Footprint Credit Scores can then be combined with the lender's in-house underwriting process as an additional variable in the model or as a matrix.

Example Digital Footprint scorecard

	Number of clients			% of total		
	Bad	Good	Total	Bad	Good	Total
8	166	524	690	24.1%	75.9%	100.0%
7	204	487	691	29.5%	70.5%	100.0%
6	224	467	691	32.4%	67.6%	100.0%
5	259	432	691	37.5%	62.5%	100.0%
4	272	418	690	39.4%	60.6%	100.0%
3	304	387	691	44.0%	56.0%	100.0%
2	353	338	691	51.1%	48.9%	100.0%
1	513	177	690	74.3%	25.7%	100.0%
TOTAL	2 295	3 230	5 525	41.5%	58.5%	100.0%

GINI	31.90%
Optimal GINI	78.90%
K-S	23.10%

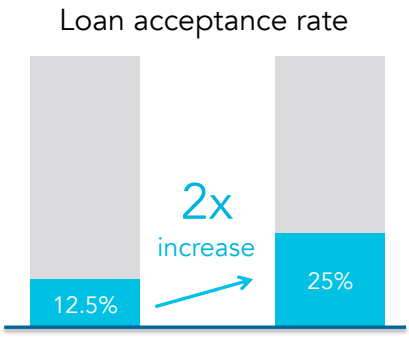
In addition to the credit scores, the lender would receive a set of Fraud Flags for each loan applicant, which helps deny loans to potentially fraudulent clients.

The credit scoring algorithms are updated periodically to adapt to changing business environment and client behavior.

CASE STUDIES

CASE 1

improvement in loan acceptance rates



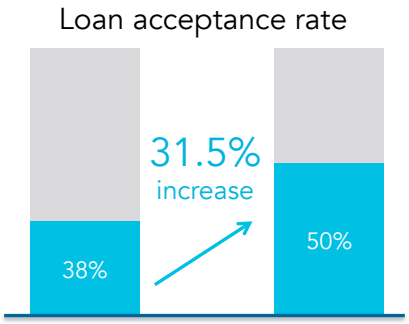
In order to keep credit losses under control, a new lender to the market was accepting just 12.5% of all incoming loan applications. This pushed up client acquisition cost and the company was struggling to reach profitability.

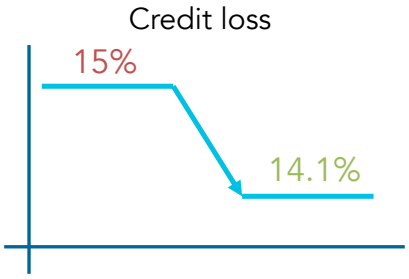
With added scoring accuracy provided by Digital Footprint, loan acceptance doubled to 25% while loan quality was kept constant.

As a result, client acquisition cost decreased significantly and the lender quickly turned profitable

CASE 2

combined improvement in loan acceptance and credit quality





A nonbank lender operating in CEE was accepting 38% of all incoming loan applications and saw on average 15% credit loss rates. Overall, the business was growing well and making money for the owners.

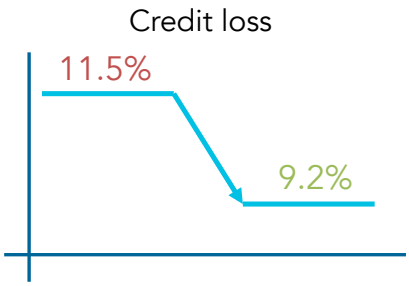
After Digital Footprint implementation, the lender was able to increase loan acceptance to 50% (31.5% improvement), which significantly reduced client acquisition cost and improved the interest and fee income.

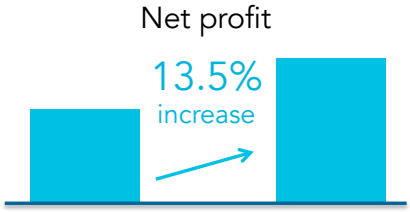
At the same time, the added scoring accuracy also allowed reducing the amount of bad loans to 14.1%.

The 2 effects resulted in both top and bottom line improvements and helped the business grow faster.

CASE 3

improvement in credit quality





The lender was happy with their loan acceptance rates and was looking for ways to further improve the credit quality.

After integrating with Big Data Scoring, the credit loss rates dropped by 20% (from 11.5% to 9.2%). Such improvement in credit quality translated to 13.5% improvement in their annual net profit.

With Digital Footprint, the additional income per each issued loan was 64 EUR. When issuing 2,500 loans a month, the added income is 160,000 EUR.

IMPLEMENTATION

Implementing the Digital Footprint solution is simple and technically takes only a couple of hours. The scores and Fraud Indicators are delivered to the lender via the Big Data Scoring cloud API and can be plugged into any lending software.

Digital Footprint implementation

1	Sign a simple contract	We will send you one
2	Send us historical dataset	We will do backtesting for model calibration. This takes a couple of weeks for new clients.
3	Install behavioural data collection tool on your lending website and connect to our cloud API	We provide the plugin and its integration is as easy as Google Analytics
4	**You're all set up! You'll get the Digital Footprint score and Fraud indicators for all your new clients.**	

STATISTICALLY SPEAKING, HOW IT WORKS?

Digital Footprint model calculates a credit score for each individual contemplating to enter into a credit relationship with the lender using our proprietary machine learning algorithms. A credit score is the representation of the default likelihood of such individual. This results in improved predictive power especially in new clients and thin credit file borrowers.

Our solutions gather up to 5,000 additional data points per each borrower from data sources that you haven't used in credit scoring before. This adds significant predictive power to the in-house underwriting process. In a nutshell, the data comes from the following sources:

Device based information

Digital Footprint takes into account in-depth information about a user's device, location and internet service provider. When a user visits your lending page, our tools capture high volume of data about which browser is being used, its version number, and details about the system, such as operating system and version. This data gets enriched by our algorithms using various databases and online search engines. All of that results in tens of parameters that are utilised for credit scoring purposes.

Behaviour based information

User behaviour on the webpage provides very valuable information about user's future payment behaviour. Since web behaviour is based on a person's subconscious decisions, it is almost impossible to fake. Moreover, while most of the data traditionally used for credit scoring purposes is based on person's historical decisions (such as credit history, income, expenses, etc.), user behaviour patterns follow the present and thus have the biggest predictive power for the future. While traditional data tries to explain one's ability to repay a loan, behavioural and other untraditional data sources tell a story about the borrower's willingness to service the loan as agreed. Having tested hundreds of possible behavioural parameters, we have identified the most valuable ones for credit scoring purposes and packaged those to the Digital Footprint product.

Web Search

Digital Footprint uses the leading web search engines, such as Google, Yahoo! Search and Microsoft Bing. This combination results in a comprehensive overview of an end user's presence in the internet. We have found strong correlation between a person's web presence and credit behaviour. In many cases, even rather simple indicators reflect it, such as how much an e-mail address is used on public websites. In addition, the sheer existence of various social media profiles in LinkedIn, Facebook or Twitter might be relevant for credit scoring.

Location based information

Using new data sources and analytical methods enables us to gain deeper insights of the specific area where the borrowers live or apply for the loan. We have found strong correlations with payment behaviour and the characteristics of the neighbourhood and the people living there. The location based
assessment uses the applicant's home address and device location as inputs and combines several various databases including but not limited to the following;

- Several economic indicators on the smallest possible scale (from county down to district level). These indicators include different salary related figures, unemployment information, etc.,
- Local databases with indicators on crime, housing prices, education, etc.,
- Points of interest, for example proximity to school and the closest park bench,
- Everything else from voting turnout to local weather.

Additional data

Digital Footprint takes into account a number of other variables. For example, telephone number given in the loan application can be enriched with the name of the service provider that usually has a strong correlation with credit behaviour. In addition, using the name of the employer for additional research using local databases like business registry results in valuable information for credit scoring.

FRAUD DETECTION

Fraudsters behave quite differently when applying for a loan or opening a credit card account based on fake or stolen identities. For example, their behaviour patterns are distinctive as they have all the required information at hand and never spend time researching it. Usually they do not bother with completing optional elements, and the way they interact with specific fields can be very uncharacteristic compared to the behaviour and cognitive choices made by genuine users. Advanced fraudsters know that their devices and IP's are tracked, so they simply change them often. However, they don't change their behaviour.

The Fraud Detection tool helps fight fraudsters by raising a flag if any potentially fraudulent behaviour is detected from the end user. The tool takes into account a multitude of factors based on the end user's device, location and behaviour. Below, you fill find some of the elements which are included in the Digital Footprint solution and would result in a raised flag;

- Unusual client behaviour: copy & paste in uncommon fields, time spent on website, etc.;
- Mismatch in client information: location based data as country, city, real location, etc. mismatch with data given in application;
- Uncommon client device information: time zone, language, IP address, etc.

Please keep in mind that our product consists of over 35+ fraudulent behaviour flags and these flags indicate a possibility that a given person is fraudulent. The risk model behind our solution is designed to recognize fraudulent behaviour and adapt to emerging fraud patterns. There is no single recommended set of values to use for deciding whether to accept, reject, manually review, or submit person to complementary services for analysis.

In determining what thresholds to set, costs of lost credit, the cost of manual review, and the cost of potentially rejecting good clients should be considered. We suggest first paying more attention to the fraud flags and once there's sufficient amount of data available, a more thorough analysis can be done to spot the most relevant fraud indicators in your business.