BIG DATA
SCORING

Digital Footprint

DATA

In-depth product overview

Digital Footprint
DATA

CONTENTS

Digital Footprint
DATA

PRODUCT OVERVIEW

Every day, lenders get some credit decisions wrong. Those incorrect decisions usually fall into 3 distinct groups;

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 number of such mistakes can be reduced with having more data about the people that apply for credit. In other words, wrong credit decisions are usually caused by not having sufficient amounts of data to make an informed credit decision.

From our experience of developing big data based credit scoring models for some of the biggest lenders on the market, we know there are vast amounts of information available from various online channels that is valuable in credit underwriting and helps make more precise and hence more profitable credit decisions.

Based on our industry leading Big Data Scoring solution, we are now launching a product for all lenders on the market – Digital Footprint Data. In a nutshell, the product provides additional information about customers that can be easily integrated to any credit scoring model. The product also includes an advanced fraud detection solution that flags potentially fraudulent behaviour and helps combat various types of fraud.

Running on the same risk models as our Big Data Scoring solution, the Digital Footprint Data is constantly updated to capture more data, recognize fraudulent behaviour and adapt to emerging fraud patterns. New functionality and parameters are added to the product on monthly basis and rolled out to our clients automatically.

Valuable credit information might come from unexpected sources. We have tested and validated data sources for many years and as a result, the product lies on the following pillars;

1	**Device based information**	in-depth information about a user's device, location and internet service provider
2	**Behaviour based information**	data based on a user's behaviour, which is mostly subconscious and hence almost impossible to fake
3	**Web search based information**	comprehensive overview of a user's presence in the internet using the various search engines
4	**Location assessment**	The Location assessment algorithm gives a comprehensive description of the consumer location as a classification cluster
5	**Fraud detection**	the fraud detection tool takes into account multitude of factors based on the user's device, location and behaviour

IMPLEMENTATION

Implementing the Digital Footprint Data product is simple and usually only takes a couple of hours. Here's the process;

1	Sign a simple contract	We will send you one
2	Fill in documents for syncing your loan application with Digital Footprint Data	We will provide the documents and instructions (e.g. last name on the loan application is labelled the same as in our system)
3	Install behavioural data collection tool on your lending website	We provide the plugin and its integration is as easy as Google Analytics
4	Connect to Big Data Scoring cloud API	We will send you instructions for that
5	**You're all set up! You'll get the Digital Footprint Data for all your new clients.**	

After the technical setup, we suggest monitoring the additional data for new clients for a few months and comparing it to the client payment behaviour. After correlations start to emerge, we suggest adding those data points to the underwriting models for more accurate credit decisions.

DEVICE BASED INFORMATION

Digital Footprint Data provides in-depth information about a user's device, location and internet service provider. When a user visits your lending page, our tools capture a 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 Digital Footprint Data algorithms using various databases and online search engines. All of that results in 18 parameters that can be easily utilised for credit scoring purposes and are part of the Digital Footprint Data output.

Table 1. Device based information parameters

Parameter	Example value	Description
Address type	Public	IP Address type: Public or Private
Latitude	51.4922	Latitude in Geographic Coordinate System (GPS). With given 4 decimal places its accuracy is +-10 meters.
Longitude	-0.02483	Longitude in Geographic Coordinate System (GPS). With given 4 decimal places its accuracy is +-10 meters.
Country	GB	IP address location based on a ISO code (e.g. GB is United Kingdom)
City	London	Name of the city based on the IP address
Postcode	E14 3SP	Postcode based on the IP address
Organization	Private	Name of the organization responsible for the IP
Proxy	No	Yes/no if the user is using proxy server
ISP	Vodafone	The name of the internet service provider (ISP)
User type	Residential	User type based on the ISP information
Device type	Desktop	The type of the device (e.g. Mobile Device, Tablet, Desktop, FonePad, etc.)
Platform	Mac OS X	Type of operating platform used by the end user
Platform version	10	Windows release version
Platform Bits	64	Platform architecture in bits
Browser	Safari	The name of web browser used
Browser type	Browser	The type of the browser (Browser, email client, etc.)
Browser version	47	The exact version of the browser
RSS	No	Yes/no if the browser is a RSS reader

Some examples for using the parameters found in table 1 above;

1. Having the latest operating system with an up to date web browser might be an indicator of better credit behaviour;

2. Using non-standard web browsers (Firefox, Chrome, etc.) might suggest better credit behaviour as the person is probably tech savvy;
3. Having a rather expensive internet service provider (ISP), e.g. Vodafone, speaks about the user's spending habits and also creditworthiness.

The above device and internet service parameters can also be used in combination with all other parameters. The best combinations for credit scoring purposes depend heavily on the specific region, loan product and customer segment. Hence, we urge our clients to test, experiment with their data and find the best solutions.

BEHAVIOUR BASED INFORMATION

User behaviour on the webpage provides very valuable information about a 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 Data product. The list of behavioural indicators that you will receive are listed on Table 2

Table 2. User behaviour parameters and descriptions

Parameter	Example value	Description
Time on website	316	Total time spent on website in seconds
Subpages visited	14	Number of pages visited within the website
Typing speed	86	The typing speed measured in keystrokes per minute
Copy & paste	2	The number of fields copy & pasted
Terms & conditions	Yes	Did the person read terms and conditions
Field changes	3	The number of fields edited after initial input

Some examples for using the parameters found in table 2 above;

1. Fewer field changes might indicate that the person is deliberate and hence trustworthy;
2. Not reading the terms & conditions usually hints that the person is less creditworthy;
3. Minimum usage of copy and paste can be seen as a sign of good credit behaviour.

The above behaviour based parameters can also be used in combination with all other parameters. The best combinations for credit scoring purposes depend heavily on the specific region, loan product and customer segment. Hence, we urge our clients to test, experiment with their data and find the best solutions.

WEB SEARCH BASED INFORMATION

Digital Footprint Data 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.

As part of the Digital Footprint Data, we will provide you the following results based on the web search about your clients. Those could be integrated to the credit scoring models or even used in marketing.

Table 3. Web search based information

Parameter	Example value	Description
E-mail	43	The amount of unique matches for the e-mail address
E-mail & company	21	The amount of unique matches for the e-mail address in combination with employer's company name
LinkedIn	Yes	Does a LinkedIn profile exist
Facebook	Yes	Does a Facebook profile exist
Twitter	No	Does a Twitter profile exist

Some examples for using the parameters found in table 3 above;

1. Having too many personal e-mail addresses present on the internet might hint that the person is less creditworthy;
2. Having social media accounts (especially LinkedIn) could be a sign of better credit behaviour.

The above web search based parameters can also be used in combination with all other parameters. The best combinations for credit scoring purposes depend heavily on the specific region, loan product and customer segment. Hence, we urge our clients to test, experiment with their data and find the best solutions.

LOCATION ASSESSMENT

Knowing where your best and worst clients originate from is an irreplaceable piece of information for great credit decisions. The Location Assessment algorithm gives a comprehensive description of the consumer location. Packed into a scientifically developed classification cluster, it can be directly plugged into your scoring model.

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. There is strong correlation between payment behaviour and the characteristics of the neighbourhood and the people living there. The Location Assessment uses the applicant's home address and device location as inputs and then combines that with info from various databases, including but not limited to the following:

- Economic indicators on the smallest possible scale (from county down to postcode). These indicators include different salary related figures, unemployment information, economic activity rates, etc.,
- Local databases with indicators on crime, housing prices, education, voting turnout, local weather, etc.,
- Points of interest, for example proximity to school and the closest park bench, number of supermarkets and bus stops, etc.

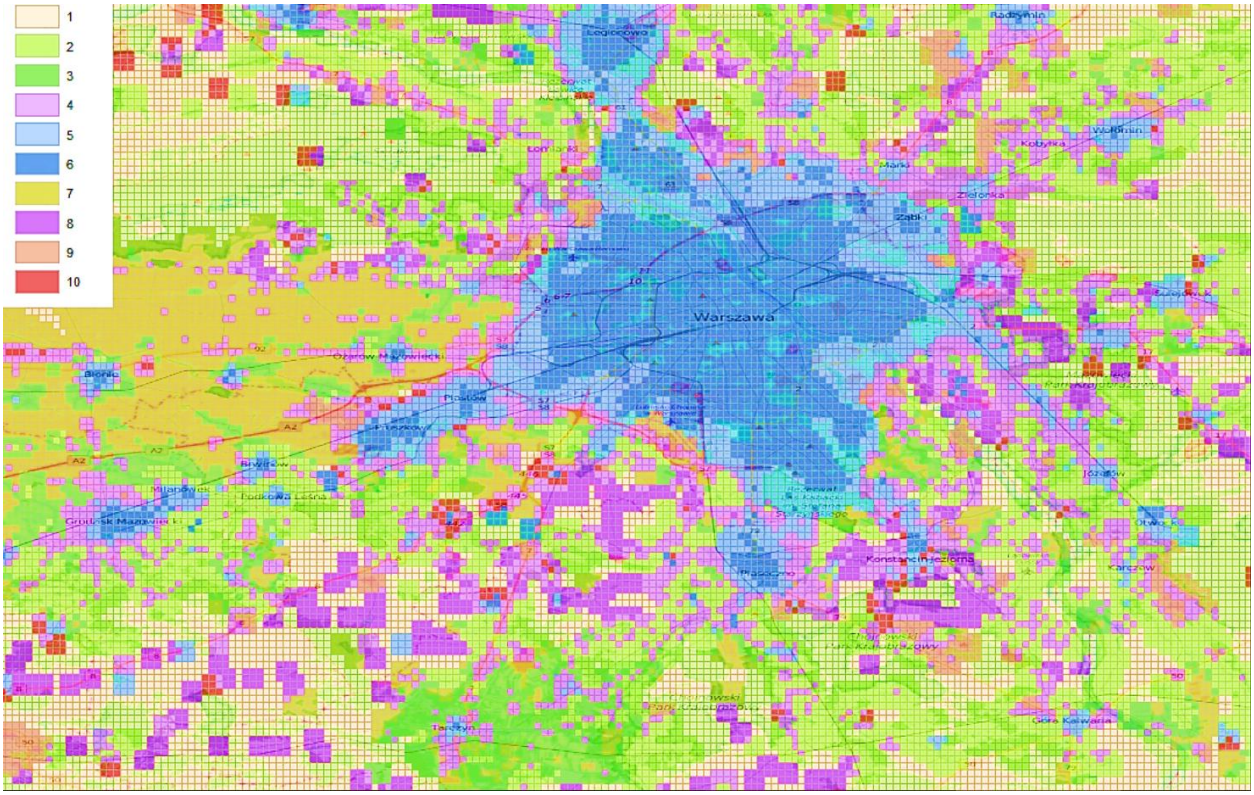
For the ease of use, we have created 10 location classifications, described in the table 4 below. These classifications can be directly used in credit scoring models depending on the specific product and target customer segment.

Table 4. Location assessment area classifications

1	Sparsely populated areas	6	City centres
2	Residential areas	7	Farmlands
3	Agricultural areas	8	Transportation hubs
4	Industrial centres	9	Industrial areas
5	Suburbs	10	Commercial centres

We recommend testing where your best and worst clients come from and then use this information in credit models and marketing (e.g. where to put billboards).

Figure 1. Example of the area classifications in Warsaw, Poland



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 Data 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.

CONTACT

Kersten Eero
Account Manager
One Canada Square (Level 39), Canary Wharf, London, UK
kersten.eero@bigdatascoring.com
+44 7535476647