



# Customer Success Dashboard GoDaddy Business Score

Baburao Kamble; Ryan Kleck; Robert Brown

# Agenda

GoDaddy's Mission for Small Business Success

Customer Success Dashboard

Machine Learning on 64M Websites Content using Spark 2.1

Business Score

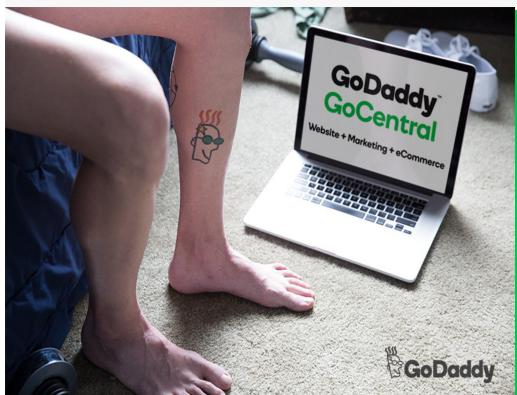
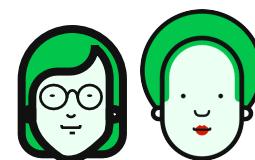
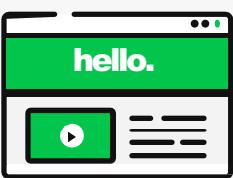
Marketing Use Case



See the Internet  
like never before.



Everything you  
need to know to  
get your business  
online.



**GoDaddy**™



**Jefferson Shallenberger**  
Sugar Percussion  
Santa Cruz, CA USA

**OUR VISION**  
is to radically shift the global economy toward life-fulfilling independent ventures.



**OUR MISSION**  
is to help our customers kick ass by giving them the tools, insights and the people to transform their ideas and personal initiative into success, however they measure it.

GoDaddy

A black and white photograph of a man in a baseball cap and polo shirt petting a golden retriever. They are in a room filled with wooden barrels, likely a distillery or brewery. The man is leaning over the dog, who is sitting and looking up at him. In the background, there are stacks of barrels and some industrial equipment.

# Product + Marketing + Care = Customer Experience

**Juniper Venus**  
Chief Comfort Officer

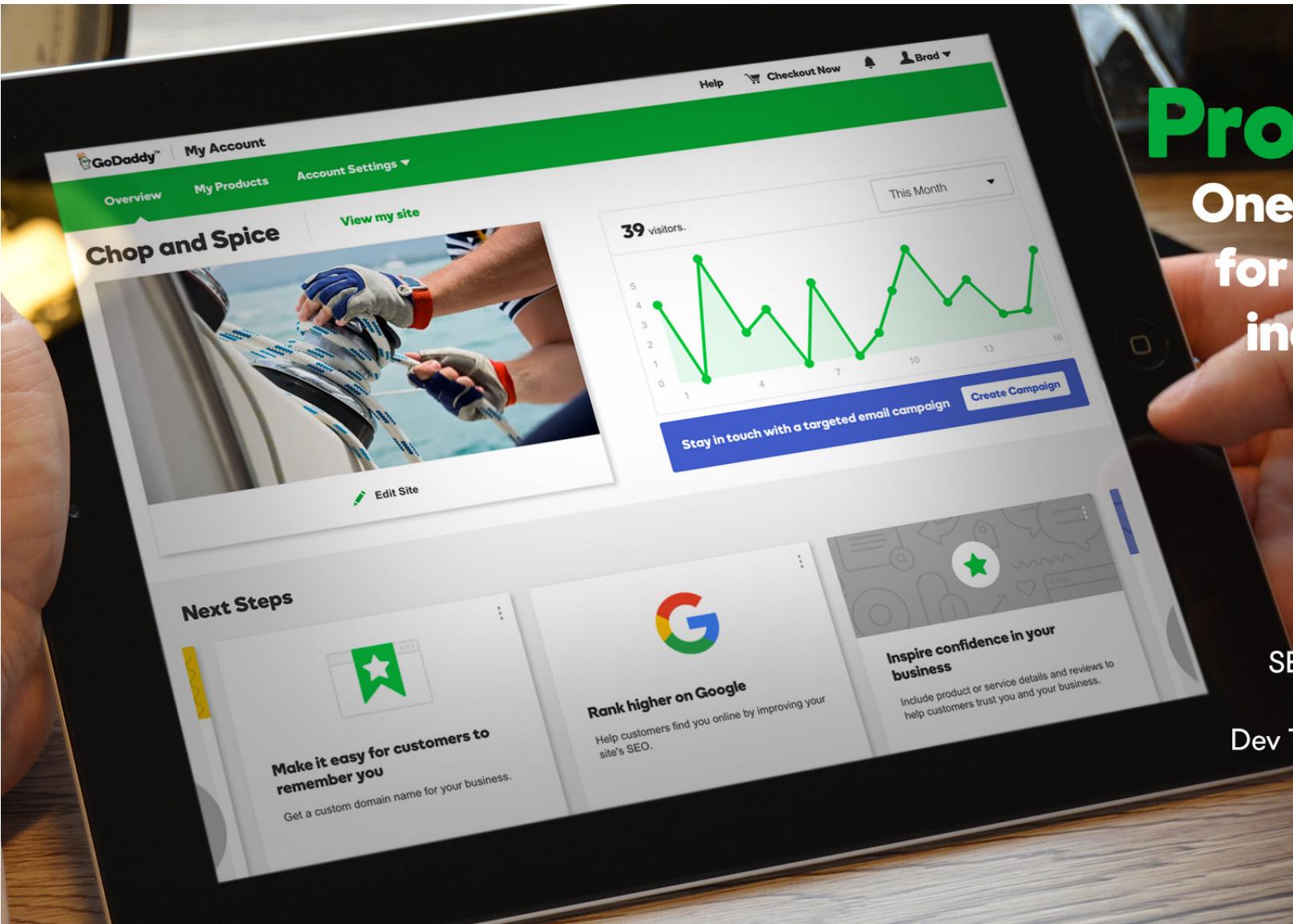
**Sean Venus**  
Venus Spirits  
Santa Cruz, CA USA

GoDaddy

# Products

## One-stop shop for successful independent ventures

Domain Names  
Website Builder  
Professional Email  
Email Marketing  
SSL & Web Security  
SEO, Listings & Social  
Business Insights  
Dev Tools & Elite Hosting



GoDaddy®

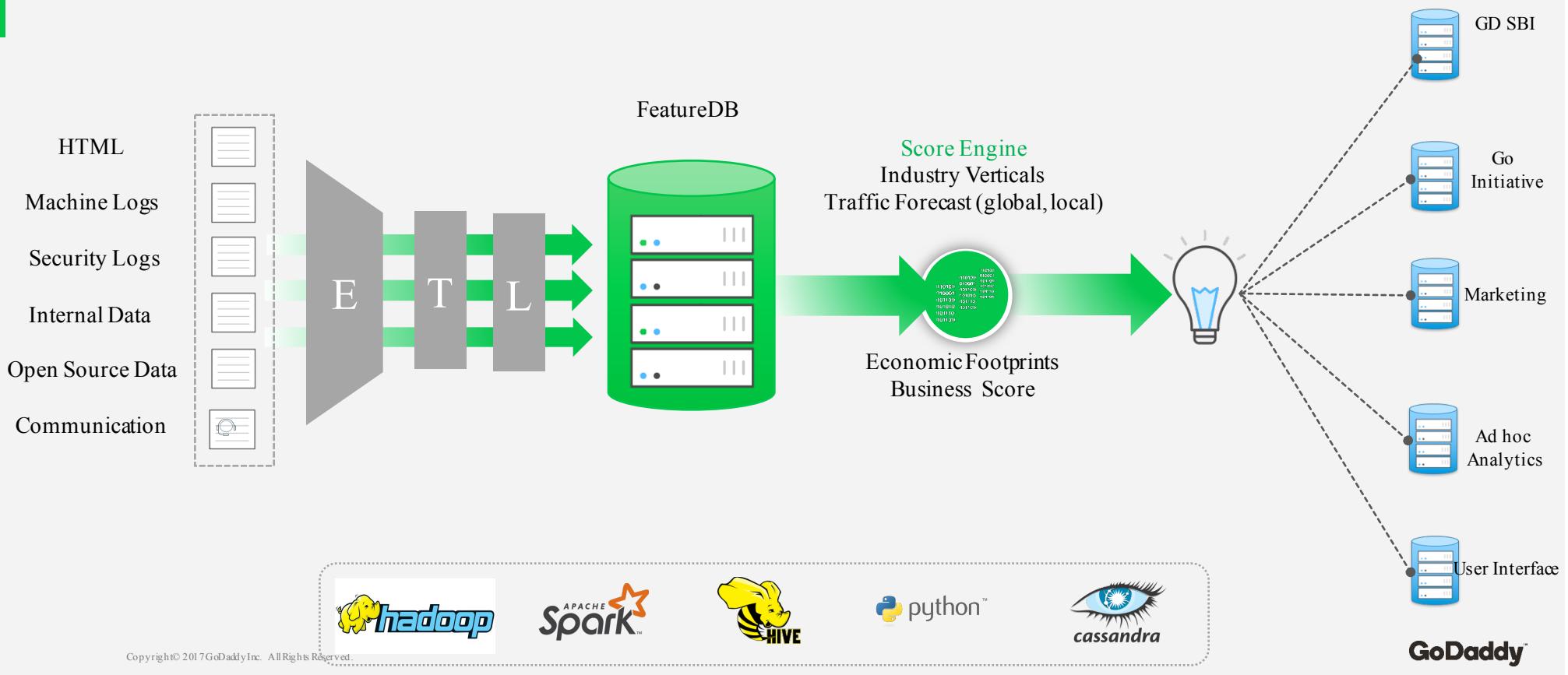


# Customer Success Dashboard

Develop the Content Analytics Platform (online and offline) for the GoDaddy's Customers to Understand Digital Presence of the Small Business in a particular industry segment.



# Customer Success Dashboard



# Components

Using Customer's website content, develop a Customer Success Dashboard from digital presence & local econometrics in the city/neighborhood where business is situated .

- Customer Success Dashboard will have all the analytics of the website done in real-time and batch according to the use case.
- Batch Processing: We are using GD data crawler as primary source of the data third party social & business API's. Use case: Customers Online Presence Index as Business Score.
- Real-time Processing: Scrape the content of the website and plug DNS address for various internal tools to create features in real-time. Use case: What you should add in your website to make small business successful.

**SEO  
site content &  
structure**

**HOSTING  
faster traffic for  
your contents**

**WEB  
website  
perfection**

**ECOM  
small business  
websites**

**SOCIAL  
social media  
coverage**

**ECON  
local economy  
index**

**INDUSTRY  
industry profiles  
by objectives**

**ECOM  
small business  
websites**

# Industry Segments/Profiles

Using Customer's website content, develop a Customer Success Dashboard from digital presence & local econometrics in the city/neighborhood where business is situated .

- Customer Success Dashboard will have all the analytics of the website done in real-time and batch according to the use case.
- Batch Processing: We are using GD data crawler as primary source of the data third party social & business API's. Use case: Customers Online Presence Index as Business Score.
- Real-time Processing: Scrape the content of the website and plug DNS address for various internal tools to create features in real-time. Use case: What you should add in your website to make small business successful.

**TLDs  
nlp on TLDs  
keywords**

**CONTENT  
naics  
classification**

**SURVEY  
self  
identification**

**PARTNERSHIPS  
sic industry  
codes**

# TLD's Verticalization

## Business Need

- Know the domains and purpose of the domain in detail by natural language processing (NLP).
- Are they personal/blog/portfolio or are they business/LLC/consultancy.
- Classify domains into 6 classes identified by modelers and marketers.
- Enable domain specific solution to enable hidden business.

## Model

- We have used internal tokenizer to tokenize domains
- These tokens feed to feature generation model which detect language, human names, & geography.
- Machine Learning model is trained on the features and keywords to classify the tld's into 6 classes.
- In 2<sup>nd</sup> iteration, business keyword rule model is designed to overwrite some misclassified data.

# TLD's Verticalization

We use Python internal tokenize and NLTK to generate the features.

containsenglish	domainname	hasname	isallenglish	keywords	language	pos	wordcount
TRUE	kidsmedicinestickers.com	FALSE	TRUE	kids medicine stickers	english	N V-N V-N	3
TRUE	actionwindowtint.com	FALSE	TRUE	action window tint	english	N V-N-N V	3
TRUE	aubihardware.com	TRUE	FALSE	aubi hardware	english	N-N	2
TRUE	michiganfellowship.com	TRUE	TRUE	michigan fellowship	english	N-N	2
FALSE	devinlamontagne.com	TRUE	FALSE	devin lamontagne	unknown	V-N	2

Above features and keywords have used to classify domains in 6 classes based on the topics modeling some of the classes are overwritten by business rules.

domainname	Segment
kidsmedicinestickers.com	Business
actionwindowtint.com	Business
aubihardware.com	Business
michiganfellowship.com	Government
devinlamontagne.com	Personal

# Industry Verticalization

## Business Need

- Identify the North American Industry Classification System (NAICS) based industry verticals for the GoDaddy domains (64M) based on the content analysis and domain names.
- Rank each domain into industry vertical based on probabilities.
- Enable business specific marketing solution.

## NAICS

- A numerical, hierarchical classification system that groups all business establishments into industries based on production process.
- The foundation on which economic census data is collected, tabulated, analyzed, and disseminated.
- A common language used among all North America.

# LDA: Latent Dirichlet Allocation

- We have developed training data using business keywords, NAICS description and keywords from the US Census data.
- We used a technique called Latent Dirichlet Allocation to classify the fingerprint of the “unknown” domain to the lineup of known NAICS Sector fingerprints.
- LDA is a generative statistical model that allows sets of observations in a linguistic corpus to be explained by unobserved groups.
- Among these algorithms, Latent Dirichlet Allocation (LDA), technique based in Bayesian Modeling, is the most commonly used techniques for nowadays.
- Assumptions on all variables:
  - Word: the basic unit of discrete data
  - Document: a collection of words (exchangeability assumption)
  - Corpus: a collection of documents
  - Topic (hidden): a distribution over words & the number of topics  $K$  is known.

# Mathematical Representation of LDA

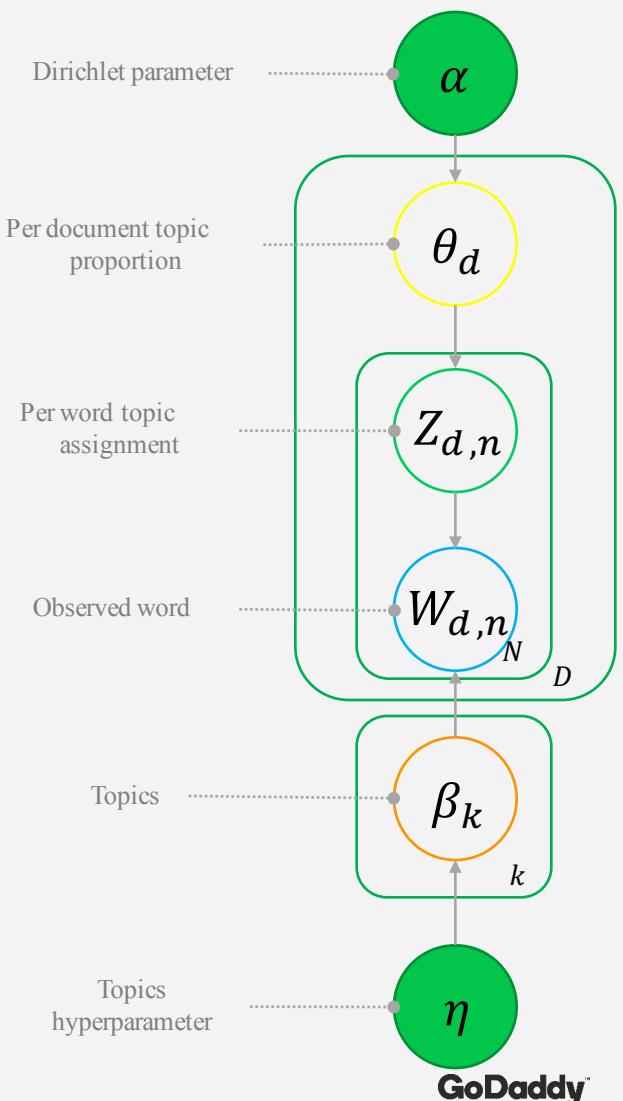
For each topic  $k$ , draw a multinomial over words  $\beta_k \sim Dir(\eta)$

For each document  $d$ ,

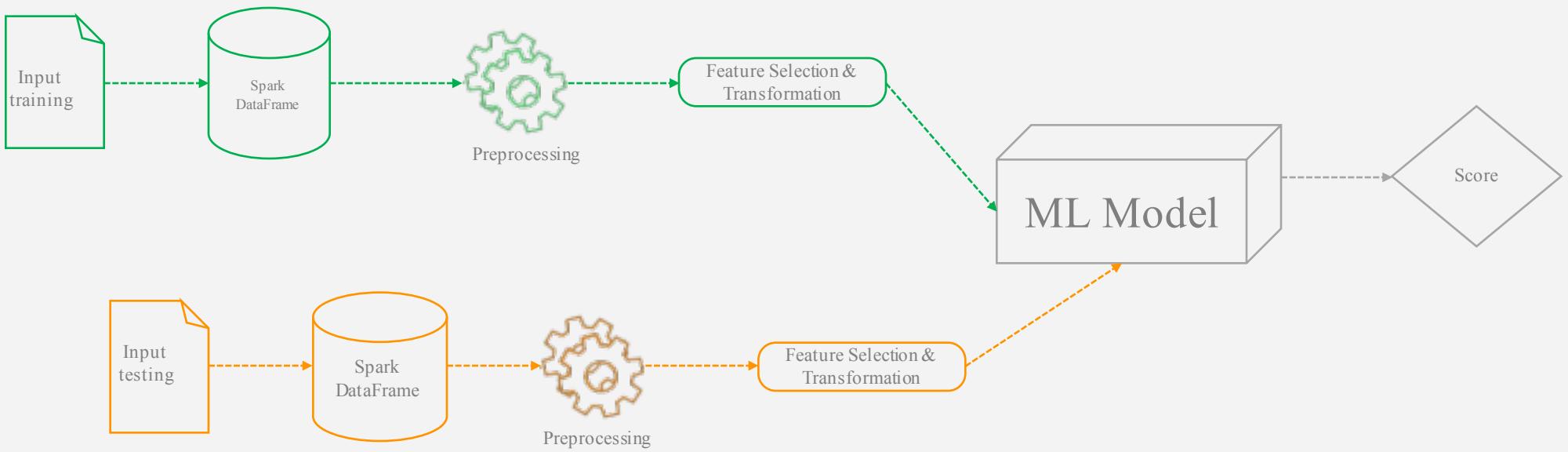
- Draw a document topic proportion  $\theta_d \sim Dir(\alpha)$
- For each word  $w_{d,n}$ :
  - Draw a topic  $z_{d,n} \sim Multi(\theta_d)$
  - Draw a word  $w_{d,n} \sim Multi(\beta_{z_{d,n}})$

$$(\beta_d | \eta) \sim Dir(\eta) \quad (\theta_d | \alpha) \sim Dir(\alpha) \quad Z_{d,n} \sim Multi(\theta_d) \quad W_{d,n} \sim Multi(\beta_{z_{d,n}})$$

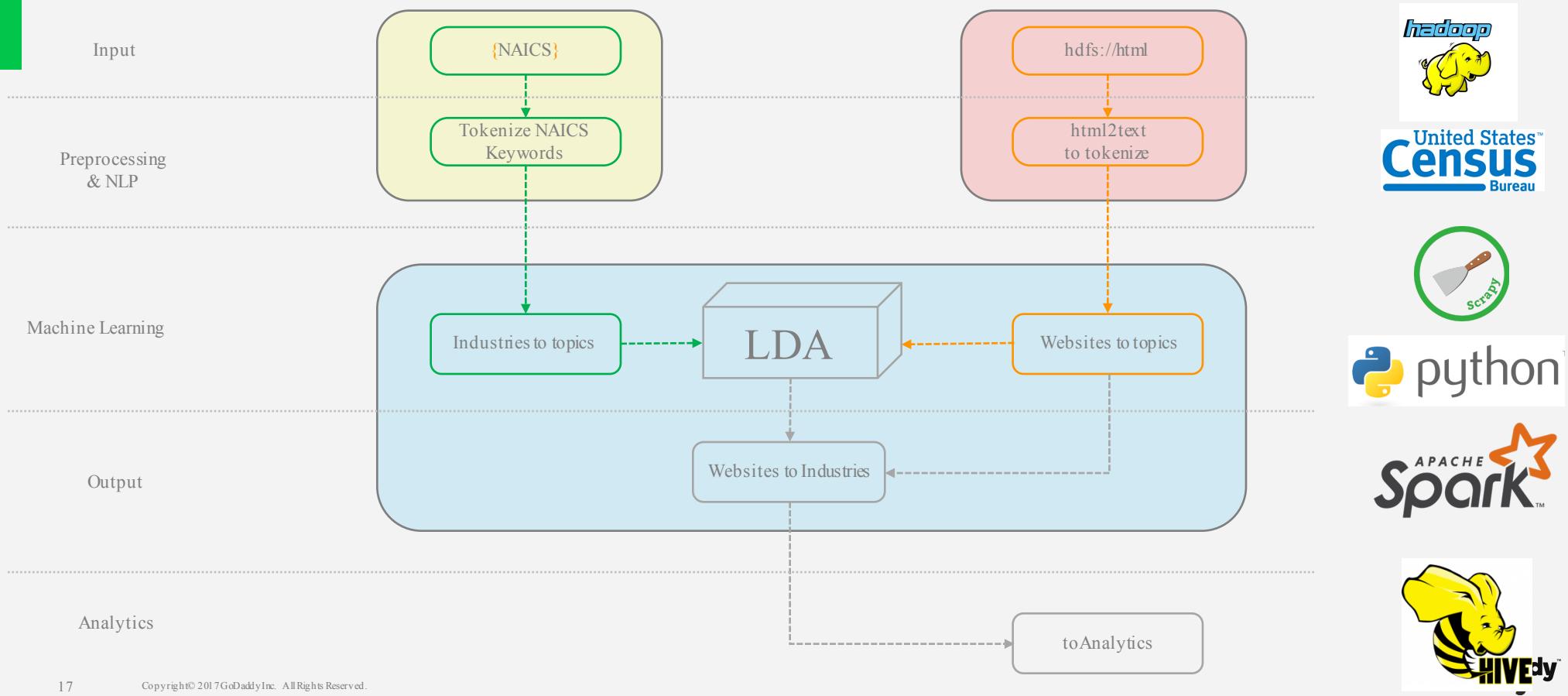
$$P(\beta, \theta, z, w) = \left( \prod_{i=1}^K P(\beta_i | \eta) \right) \left( \prod_{d=1}^D P(\theta_d | \alpha) \prod_{n=1}^N P(z_{d,n} | \theta_d) P(w_{d,n} | \beta_{z_{d,n}}, z_{d,n}) \right)$$



# Modeling Approach



# Industry Verticalization



# Sample Results

Industry Vertical Sample Score for GoDaddy.com

```
{'godaddy.com':[0.0, 0.037261519371214455, 0.12163335800668636, 0.12212609505637625, 0.0, 0.0, 0.0391883272323731, 0.006822915544477993, 0.07649356215522182, 0.03418770216006102, 0.08399637531579494, 0.0, 0.044632956375623346, 0.08008591676796364, 0.23397736516285688, 0.023100845006343924, 0.05476212126480185, 0.045210897428489265, 0.01195930071056271, 0.05362539333191509, 0.030935349109237273, 0.03476925258258562, 0.028247390216576837, 0.04214249732233799, 0.11761814220985195, 0.0, 0.03485210250149108, 0.0, 0.036624106810977664, 0.023490911619049086, 0.08161086708509405, 0.09972968094058156, 0.024987767078109006, 0.01669645821975547, 0.027306928087764512, 0.0, 0.0, 0.59144987630468255, 0.01714920401645141, 0.057044712502075504, 0.33934270898168075, 0.026937393520935108]}
```

Top three Industry verticals with probabilities

Row(domainname=u'godaddy.com, 541=u'0.591449876', 551=u'0.339342709', 518=u'0.233977365')

**518 - Data processing, hosting, and related services**

**541- Professional, scientific, and technical services**

**551- Management of Companies and Enterprise**

# WEB

- For Small business's website, it's vital to position business online with a strong, professional destination that gives customers the impression you mean business and the motivation to want to engage more with your business.
- We have created the process to analyze the website based on the What makes the Small Business Website perfect. This helps visitor connect and checkout the important section of the website easily from website and mobile.

**WEBSITE  
website  
completeness**

**METADATA  
Analyze metadata  
completeness.**

**CODING  
coding style &  
text to code ratio.**

**DESIGN  
modern UI/UX  
features templets**

# ECOM

- For online small ecommerce business, it's vital to make the secure and professional website to sale products online.
- Based on prominent ecommerce websites, we have created the important component for successful ecommerce business and checked these components on the websites.
- We have created the process to analyze the website for the ecommerce components (shopping cart, payment solution, SSL, shipping, etc.) .

**SECURITY**  
**web security &**  
**SSL encryption**

**ANALYTICS**  
**analytics plugins**  
**& campaign.**

**CART**  
**online shopping &**  
**payment solution**

**CONATCT**  
**contact, chat,**  
**email, maps etc.**

# HOSTING

- Most small businesses end up making a wrong choice in the beginning because they do not properly evaluate their needs.
- We have created the process to analyze the hosting based on the content, SEO and website content size, speed and hosting information

**A-RECORDS  
hosting  
information**

**BANDWITH  
self  
identification**

**CONTENT  
type & size of the  
content**

**SPEED&COUNTS  
download speed &  
dns count month**

# SEO

- Small business's definitely needs to have an SEO strategy in place if you are interested in succeeding in terms of online marketing.
- investing in organic SEO is more important for small businesses now than ever before.
- We have created the process just like SEO experts to analyze the keywords of website's meta title, meta description, h1, h2, and image alt tag. This helps search engines connect with website with the keyword, making web page more relevant.

**METADATA  
metadata  
information**

**WEB  
web elements &  
keywords**

**ROBOTS  
check robots.txt  
& sitemap**

**LINKS  
analyze all urls &  
links on website**

# ECON

- Small business's growth is totally depends on the local economy and purchasing power of the town.
- Local economy has big impact on the Small Businesses and vice a versa.
- We have used online open dataset like tax, income, housing prices to create the ECON index.

**INCOME**  
**average income of  
the family**

**TAX**  
**local small  
business tax**

**HOUSING**  
**housing price  
index**

**CITY DYNAMICS**  
**population,  
growth, tourism**

# SOCIAL

- For Small business's there is need attract customers to local shops or online ecommerce business with a strong online presence.
- Traffic to the website or local store can be driven through the large social media efforts and listing in review websites like yelp, foursquare etc.
- We are using third party api, data providers and listings to aggregate the score as the index of popularity of business.

**SOCIAL  
social media for  
small business.**

**LISTINGS  
local business  
listing (e.g. yelp)**

**MEDIA  
Web elements &  
keywords**

**LINKS  
Analyze all urls &  
links on website**

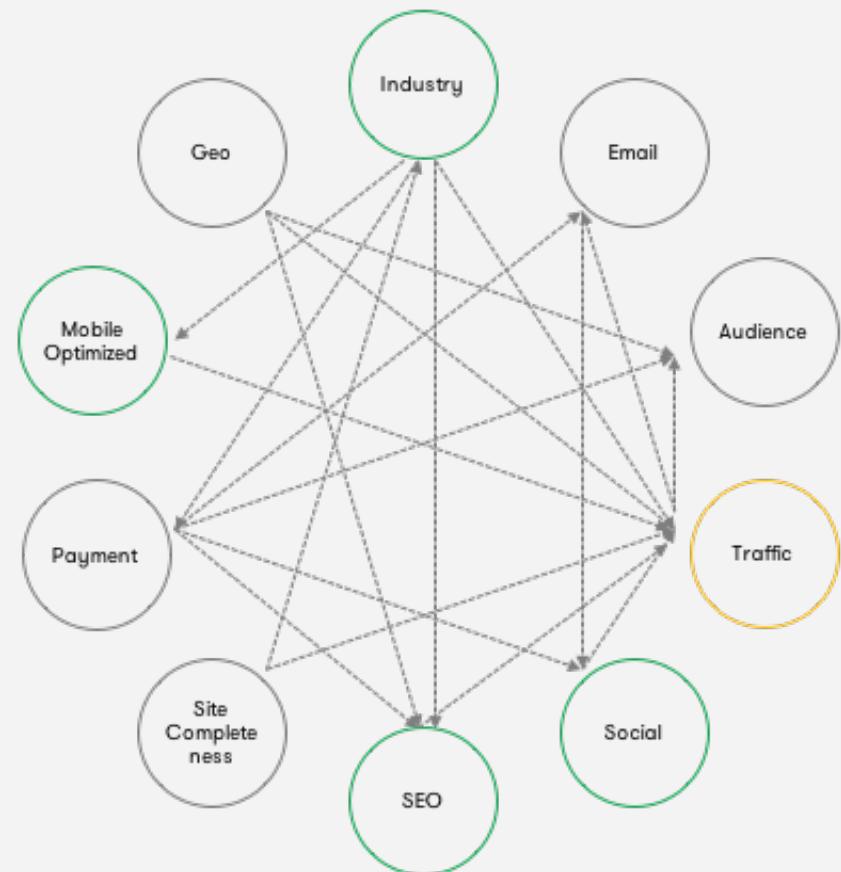
# GoDaddy Business Score

- GoDaddy Small Business Index (score) is a digital number derived from a mathematical interpretation through complex machine learning process on the digital information derived from the complex data.
- Our Score is a measure of how well website is presented, secured, socialized and how well website is doing relative to other sites in same industry segment.
- We use the Bayesian Network to select the features and Gaussian constants from the feature database.

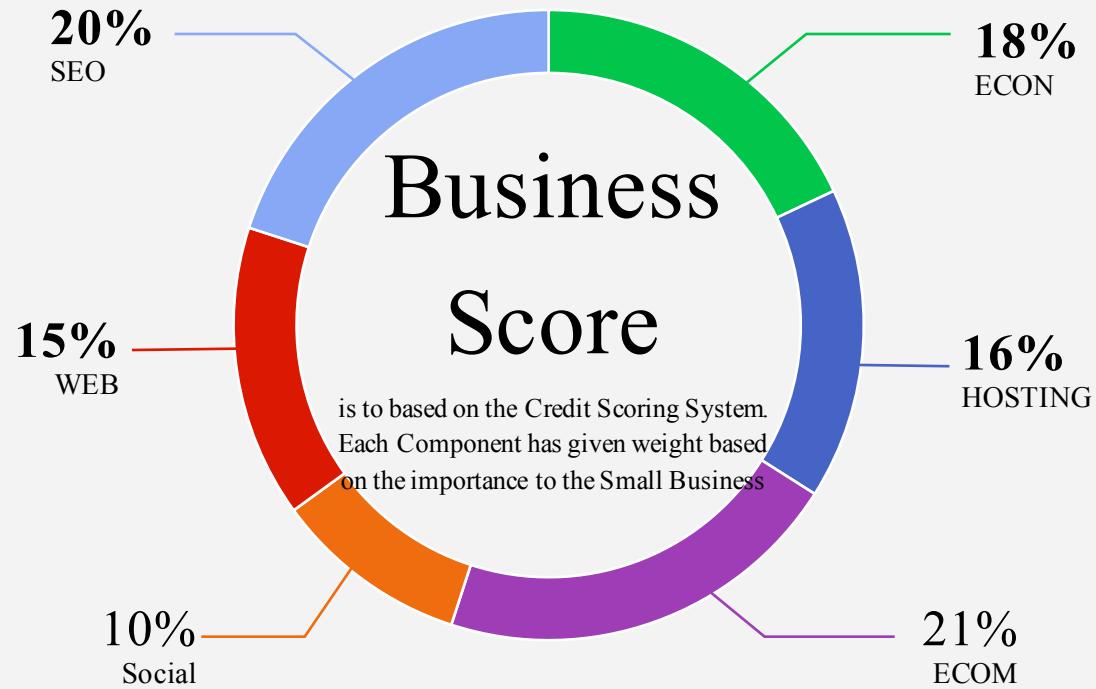
# Web Features Relevancy by Industry Segments

Bayesian network diagram shows the simple relationship between the different components (data) of the Customer Success Dashboard.

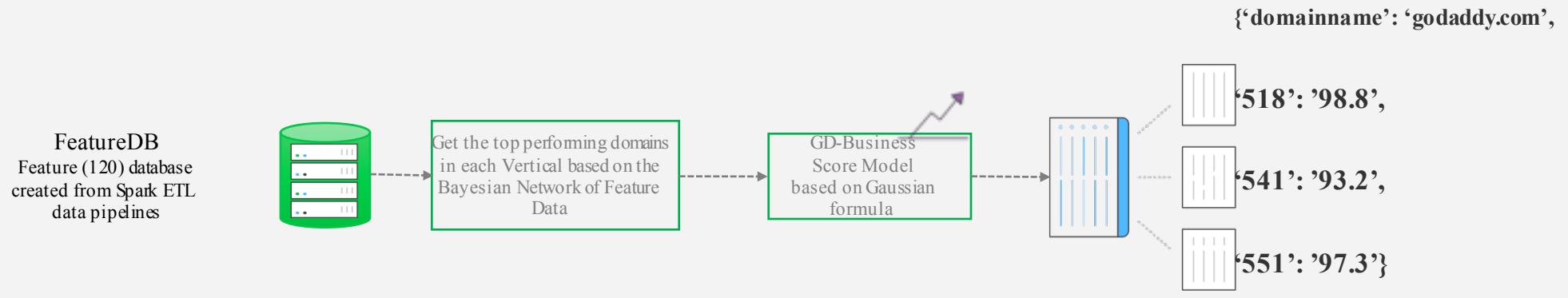
For example one can see that SEO, industry, social (Facebook, twitter, etc.), mobile optimized and site completeness all have causal effects on the likelihood of the website traffic as expected. The relationship between the industry and traffic also becomes clearer.



## Business Score



# GoDaddy Small Business Index



GoDaddy Business Score/Index is similar to the credit scoring system for digital presence.

$$GoDaddySBI = f \left( \int_0^{100} f * Score(Hosting, Communication, Industry, Engagement, Audience, Econometrics) * x_f \right)$$

Each of digital component has given weight.

We analyze and rescale the score based on the features of the website with top 100 domains in the each industry segment.



# Marketing Use Cases

# Mobile Friendly Websites

## MISSION

### Increase Mobile device based SEO

- Let's make sure GoDaddy Customer's website shows up in search results.
- Detect the whether website is configured for multiple devices & help search engines understand Small Business website to show in mobile device based search.

## MODEL

### Text Analytics on Raw HTML data

- We have designed the algorithm to detect the responsive web design which is recommended design pattern by major search engines.
- Algorithm also detect mobile optimized websites.
- Algorithm is validated with Google's Mobile Friendly tool with 0.2% sample which reported 94% accuracy.

## REVENUE

### Help to Small Business.

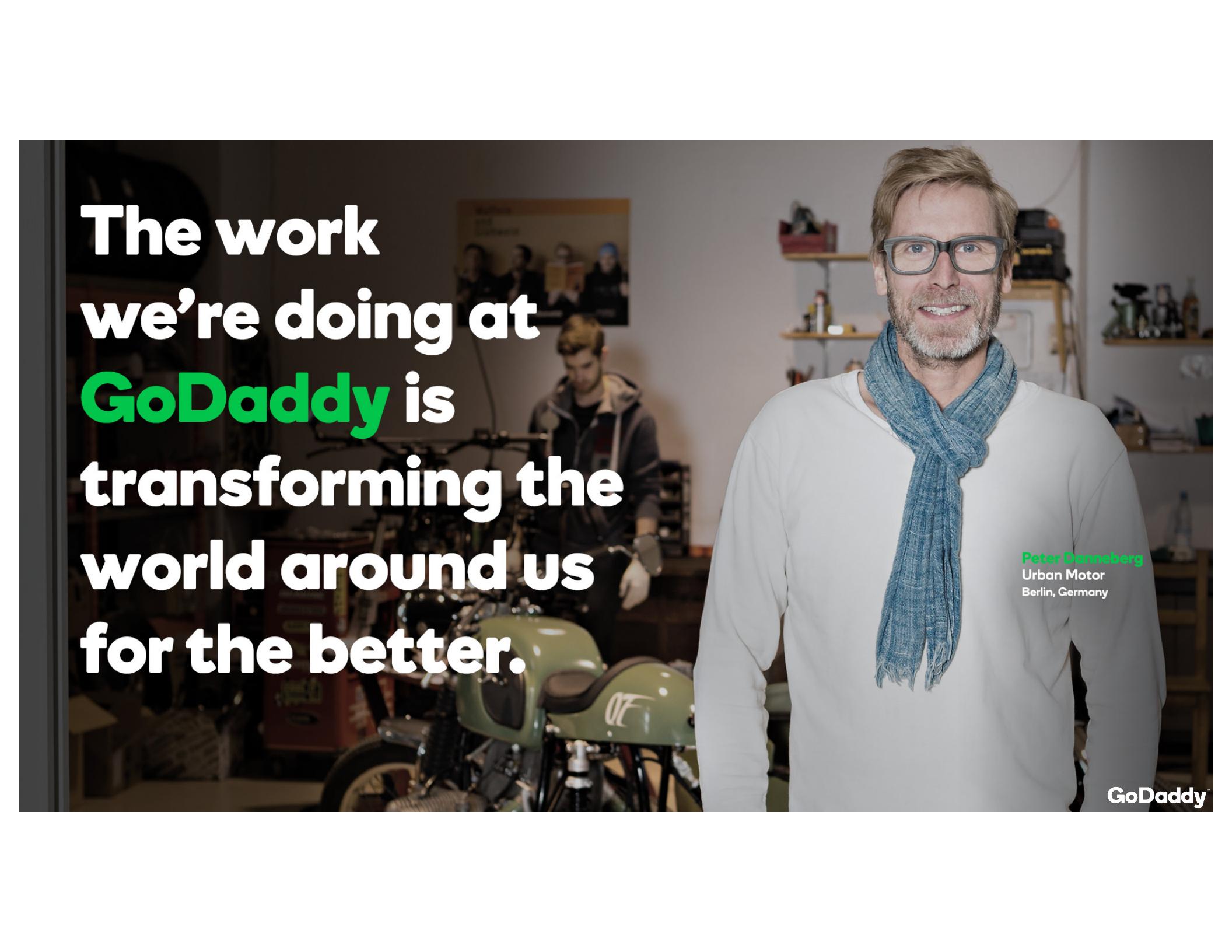
- We send notification to the shopper and advice them based on the outcome of the mobile friendliness of the websites.
- GoDaddy also help shoppers to make websites mobile friendly on all devices.

## Cart + Payment - SSL

MISSION	MODEL	REVENUE
Secure the transactions	Text Analytics on Raw HTML data	Help to Ecommerce Small Business.
<ul style="list-style-type: none"><li>Let's make sure GoDaddy Customer's online small businesses are safely accept the credit cards.</li><li>Detect the whether website has shopping cart, payment solution and SSL.</li></ul>	<ul style="list-style-type: none"><li>We have designed the algorithm to detect the SSL, it's type, shopping cart and payment solution provider.</li><li>Algorithm also SSL detect encryption type.</li></ul>	<ul style="list-style-type: none"><li>We send advice shoppers based on the outcome of the features detected from the websites.</li><li>GoDaddy also help shoppers to make websites more secure and shopper friendly to accept online transactions.</li></ul>



GoDaddy's Business Score goal is to help established small businesses get to the next level by increasing online presence.



The work  
we're doing at  
**GoDaddy** is  
transforming the  
world around us  
for the better.

Peter Danneberg  
Urban Motor  
Berlin, Germany

GoDaddy®