### Embracing Big Data

CMU Data Analytics Conference | September 22, 2017

Lonnie Miller Principal Industry Consultant, SAS



Copyright © SAS Institute Inc. All rights reserved.







# WARGAMING.NET

### Embracing Big Data – Today's Discussion



### What Business Leaders Care About

Data, Technology & Effective Processes





#### Relevant Skills & Capabilities



## used by **91** of top **100** Fortune Global 500 companies

in 148 countries

at 80,000 sites





### What Business Leaders Care About



## lanage 11 A 1-percentage-point increase in customer loyalty translates into an additional \$700 million in revenue for GM <sup>77</sup> Automotive News, 6/10/13 Jstome





GM working on connected and automated vehicles in Michigan

0 💿 💟 🛅 😂

Michigan road agencies and General Motors R&D division are collaborating in the connected and automated vehicle



Move over connected cars (Reality Check)

0 💿 💙 🛅 😂

4 reasons IoT on wheels is accelerating and will be bigger than the connected car market With headline-grabbing bells and whistles, the growing cult around connected cars is understandable. The booming ...



Investing in IoT: three facts behind the hype

#### 6 💿 💟 🛅 😂

By now, we've all read and heard about the internet of things. Cars that drive themselves will carry us to buildings that recognize us when we enter, and unless our ...



Analyst Angle: Analytics to optimize enterprise and IoT applications

6 💿 💟 🛅 😂



Scale, security, functionality – industrial IoT (Reality Check)





Renovating the smart home industry (Reality Check)

6 💿 💟 🛅 😂



What is the role of IoT in digital transformation?

f 💿 💟 🛅 😂

CELC

LG creates new AI and robotics research units

#### 6 6 🖸 🗇 🖸

LG said these new technologies will be applied to the firm's existing technology portfolio Korean company LG Electronics announced the establishment of two new research centers

#### Executives Don't Think About Big Data



ariety



### Executives Think About What Big Data <u>Will Do</u> for Their Business



ctions

cceleration

dvantages



### ... human behavior? How does .loyalty? analytics quality? safety? influence content



#### **Commercializing** Connected Vehicle Data

Safe driving

Influencing human behavio in app, in car

**Eco-** driving

Multi-modal user score Optimal fuel consumption

PAYD PHYD PFMD

Verbatim and text categorization

**Optimal offers** 

**Context aware** 

Fraud detection



#### **Optimizing the Customer Experience**

#### Personalization

#### Recommendations

### Understand caller intent

#### **Goodwill treatment**



Discover meaningful customer interests or concerns

Placement into profitable loyalty tiers



#### **Intelligent Customer Interactions**



## How'd those assorted tank tops work out for you?





Copyright © SAS Institute Inc. All rights reserved.

How Does Advertising Benefit from Real-Time Analytics?







#### Manufacturing Excellence

#### Preventative

#### maintenance

#### Situational repair recommendations

## Data mining for root cause investigations

### Escalate early (perceived) quality issues



**BEE**ST

#### **Better Business Planning**

Demand Planning / Sales & Parts, Contract & Delinquency Forecasting

Experiment with alternative forecast techniques Model performance (champion / challenger

assessment

Identify causal factors



- 23.86



### Data, Technology and Effective Processes



### What Is This?

"PEJCU0Q+PEJIdIJIcG9ydD48QmV2UmVwb3J0X0 IuZGV4PjxyYXdSZXF1ZXN0PjIyQTAwNTwvcmF3U mVxdWVzdD48ZWN1QWNyb255bT5CUENNPC9I"



### Raw Data from a Vehicle





### Types of Big Data You Should Care About

<b>Structured</b>	<b>Raw</b>	Machine-generated	
(e.g., numeric, fixed length)	(e.g., POS terminal)	(e.g., sensors, RFID)	
Unstructured	App logs	<b>Streaming, real-time</b>	
(e.g., text)	(e.g., sign-in times)	(e.g., cloud altitude)	
<b>Geospatial</b>	Social media	<b>Time-series</b>	
(e.g., lat/long, GPS)	(e.g., tweets, blogs)	(e.g., sales history)	
<b>Event</b>	Weblogs, clickstream	Image	
(e.g., alerts)	(e.g., mobile site visits)	(e.g., cancer cells)	



#### **Best Practices**





### What Could You Do If You Found These Customer Segments in Your Online Order Data?



#### What If ... ?

- Isolate profiles: 1. profitability, seasonal or regional uniqueness
- 2. Identify popular SKUs per group
- What's common 3. between #1 & #2?
- 4. Examine sentiment based on text, CSR notes, surveys &/or warranty records
- 5. Identified overlapping SKUs



**DEFECTION RISK** 

High

\$

\$\$\$\$



### Possible Roadmap & "Quick Wins"

#### Strategic & Tactical Marketing Options

- 1. Shipping (examine DCs) aiding most profitable segments?
- 2. Content audit (call center scripts, improved web or mobile site content changes based on sentiment or concepts via text analysis)?
- 3. Retention program for #3?
- 4. Price testing with #1?
- 5. Prioritize forecast initiative for overlapping SKUs



**S**sas

### **Technology Enablers**



### Example: On-site Service Offer Experience



5

- A reminder of other upcoming maintenance needs
- The name of the onsite service managers she may speak with to inquire about the offer



sms

#### RTDM will execute decisions to determine the best interaction:

 (Illustrative ex:) Model uses owned vehicle segment, enrollment for owner notifications, prior # service visits, new vehicle in-market score and prior customer pay amounts



Customer Lifetime Value (CLTV) score is updated

RTDM will interface to SAS Marketing Optimization (MO) to retrieve optimized offer set









ESP looks up Paige's guest details based on phone / email / user information from her Volkswagen Credit account

SAS Real Time Decision Manager

### "Goodwill" - Real-time Decisions

Long-Term Value (LTV) Model

	1	2	3	4	5
	(~\$4,999)	(\$5,000~\$9,999)	(\$10,000~\$14,999)	(\$15,000~\$29,999)	(\$30,000~)
5 (50%~)	\$	\$\$	\$\$\$	\$\$\$	\$\$\$
4 (25%~49%)	\$	\$\$	\$\$	\$\$\$	\$\$\$
3 (5% ~ 24%)	\$	\$\$	\$\$	\$\$	\$\$\$
2 (1% ~4%)	\$	\$	\$\$	\$\$	\$\$
1 (0% ~ 0.9%)	\$	\$	\$\$	\$\$	\$\$

#### GOODWILL OFFER FOR BRAND RETENTION

LOW RISK	MEDIUM	HIGH RISK
\$	\$\$	\$\$\$

Forbes magazine<sup>1</sup> identified a medical research facility generating 100 terabytes of data that was ultimately copied and retained by 18 different teams and required more than 10 petabytes of storage.



forbes.com/sites/ciocentral/2012/07/05/best-practices-for-managing-big-data

### This is Madness!



Source: The Five Essential Components of a Data Strategy, SAS 2016

#### Data Strategy



Figure 1: The five core components of a data strategy.



### Is This Acceptable?



Figure 6. Length of time to put a model into production. Based on 141 respondents who stated they are doing this today.

Source: Operationalizing and Embedding Analytics for Action, TDWI 2015





### **Relevant Skills & Capabilities**





#### **Desired Qualifications**

#### The Company:

Faraday Future is a California mobility company bringing a tech approach to clean transportation. At FF, we believe that by placing equal emphasis on automotive and technology disciplines, our team of experts is uniquely positioned to take a usercentric, technology-first approach to vehicle design with the ultimate aim of connecting the automotive experience with the rest of your life.

#### Your Role:

- TLDR: Use Stats, Machine Learning, Math and CS to help design the car and its sensors, gain insight into business decisions regarding sales & marketing, help create a seamless service experience and model reliability.
- Work in the IoV space to help develop the next generation of algorithms and methodologies that will power the connected vehicle.
- · Understand the consumer space to develop next generation customer sales and service models.
- · Utilize machine learning algorithms to help customize the car around around the user.

#### Preferred Requirements:

- · PhD in Statistics, Biostatistics, Computer Science, Applied Mathematics or related field.
- Experience with Pig or Hive.
- Experience with NLP and Sentiment Analysis.
- Prior work history in the automotive industry.
- EE or Hardware experience with sensors a plus.

 $\infty$ 

#### Data Scientist

SAN JOSE, CA DATA SCIENCE FULL-TIME



#### **Desired Qualifications**

#### Advanced Analytics Analyst

Trexin Consulting \*\*\*\*\* 2 reviews - Chicago, IL Contract

Trexin Consulting is a management and technology consulting firm specializing in the application of advanced technologies that drive business value. We help our clients devise business strategies that capitalize on disruptive changes related to technological innovation, government regulation, organizational restructuring, and business-model shifts. Our expertise is aligned with the industries that we serve, including healthcare, financial services, products & distribution, and new media / high tech, and we have national coverage with clients in Chicago, Miami, Minneapolis, New York and San Francisco. We are currently seeking a Contractor to join our team and work as an Advanced Analytics Analyst in Chicago.

#### Job Description:

This position will be considered a leading Healthcare Analytics Domain Expert and will have expertise in evaluating the value of new features for Clinical and Non-Clinical Data Acquisition, strong predictive modelling skills and experience consulting and providing critical insights for all lines of healthcare business including Commercial, Medicare and Medicaid. This position will consult with clients to develop solutions, share analytical insights, to test and drive product innovation and enhancements.

Employs advanced analytics to gain critical insights into member behavior drivers and preferences, product/program concepts and value propositions, operational effectiveness and efficiencies, client specific health gaps and needs.

Consults business leaders and internal partners to apply the insights to test and drive product innovation and enhancement, to target the right members for the right service, to measure program outcome and demonstrate value, to help win and retain clients, and to support information based strategic decision making.

#### Primary duties may include, but are not limited to :

Provides analytical insights to support business solution development based on quality, use, cost and other key drivers; develops and refines analytical plan templates for problem diagnosis and opportunities assessment; consults on all considerations related to designing and executing tests or pilot programs; contributes to the design of new program/initiatives based on test/pilot outcomes; develops requirements and targeting criteria, member segmentation, understands product design principles and operational considerations; develops predictive models and other tools that help target the right members; develops methodology to measure engagement, clinical, utilization and financial outcomes of a program/initiative; conducts in-depth research to address challenging issues in measuring outcomes; leads applied research projects; authors white papers, presents at industry conferences and submits journal publications. Addresses limitation of certain data sources and evaluates incremental value of new data sources. Requires Graduate-level degree with concentration in a quantitative discipline such as Mathematics, Statistics, Economics, Epidemiology, Engineering, Computer Science or Operations Research; 3-10 years' experience in related healthcare analytics related field; or any combination of education and experience, which would provide an equivalent background. PhD preferred. Experience with predictive model development lifecycle for segmentation/stratification that supports customized strategies for different customer segments in a major industry highly desirable - healthcare, financial, database/direct marketing, Search Analytics, Web Analytics and Online Retail. Consulting, Substantial analytical experience in healthcare industry preferred. Advanced experies with SAS, R or equivalents analytical tools comprehensive experience with Teradata. SQL, Hadoop, or equivalent data base tools.

#### Preferred Candidates should be able to:

Understand how to use healthcare industry code sets which includes healthcare knowledgebases, taxonomies, electronic medical records/health records, and medical devices. Should be able to interact with senior management to support the decision-making processes and serve as the Data Science and Healthcare Analytics Business Domain Expert on cross departmental initiatives, should be able to act as a facilitator, synthesizing feedback from medical directors and business leaders in product/concept ideation activities and develop and validate business cases for data science projects.

#### Additional skills and experience desired:

- Experience and aptitude in Data Science and Advanced Health Care Analytics including but not limited to Clinical Health Profiling, Payment Integrity Analytics, Financial Analytics, Provider Analytics, Search Analytics and Web Analytics
- · Healthcare Analytics Domain Expert
- Strong predictive modeling skills and experience consulting and providing critical insights for all lines of healthcare business
- Substantial analytical experience in healthcare industry preferred
- · Advanced expertise with SAS, R or equivalents analytical tools
- · Comprehensive experience with Teradata, SQL, Hadoop, or equivalent data base tools

#### Job Type: Contract

Required education:

Master's

Required experience:

· healthcare analytics: 1 year



#### **Desired Qualifications**

<b>Ford</b> Go Further	Automated Driving Research Scientist Machine Vision & Learning Ford Motor Company – Palo Alto, CA			
Apply on Company Site 🖄 🛇 Save				

#### Job Description

Job Description

#### Job Summary:

Ford Motor Company is moving into a new phase of its 100+ year history, one in which openness is at the heart of its future. The automobile is being redefined as a networked computing platform upon which an ever-evolving set of applications is being designed, allowing us to create exciting new consumer experiences.

We are a research lab within Ford that is investigating computing platforms, software & hardware architectures, sensors & strategies for fusing information from a variety of external & internal environmental sensors for next generation active safety and autonomous vehicle technologies.

Candidates for consideration will be responsible for supporting research and developmental efforts in automated vehicle technologies ranging from Advanced Driver Assistance Systems (ADAS) to semiand fully-autonomous driving capabilities. The ideal candidates are exceptionally creative, motivated and talented engineers with strong interdisciplinary backgrounds in science & engineering with solid programming skills. We collaborate extensively within Ford and with external partners, thus we require exceptional communication and collaborative skills.

These positions are located at Fords Research and Innovation Center in Palo Alto, CA

#### Job Responsibilities:

- Design, prototype, and evaluate effective and efficient sensor feature detection, tracking, mapping and localization algorithms and systems, including computer vision algorithms
- Design, prototype, train and evaluate various machine learning methods for use in detection, classification and regression tasks, significantly improving performance or reducing cost
- Develop, evaluate and compare deep learning algorithms for specific applications and tasks to
  improve performance, data collection, training, and suitability for embedded applications
- Participate in technical discussions and collaborate in the creation of new ideas for automotive applications within the existing autonomous vehicle research team; coordinate with engineers in related Ford R&D teams; contribute to technical roadmap for future development
- Maintain close contact to the scientific and industrial community in computer vision and machine learning, and perform scouting and assessment of new approaches
- Work closely and effectively with external partners corporations, universities and automotive suppliers
- · Publish/present technical reports, papers and/or pursue Intellectual Property rights

#### Job Requirements:

#### **Basic Qualifications:**

- Ph.D. in Engineering, Computer Science, Natural Sciences, Robotics, or Statistics
- · 3+ months hands-on experience in Deep Learning or Machine Learning
- 1+ years of experience in multiple contemporary computer programming languages, e.g. C/C++, Python, Java, OpenGL, PERL, MATLAB/Simulink, and software libraries and tools

#### Preferred Qualifications:

- 5+ years of Computer Vision and/or Machine Learning hands on work experience
- · Expert experience in designing and building state of the art feature detectors and trackers
- · Demonstrated ability to carry out independent research and lead projects
- · Experience with computer vision, machine learning, DNNs, and numerical optimization
- · Passion to develop and extend deep learning technologies
- Experience with algorithms such as statistical signal processing, image processing, simultaneous localization and mapping, geospatial location, rendering 3D data, computer graphics, etc.
- Knowledge of 3D co-ordinate frames and transformations, vector mathematics, matrix algebra, state estimation, probabilistic inference and modeling, etc.
- Experience developing and optimizing algorithms and systems with reduced computational complexity or cost
- Experience in multiple computer hardware, operating systems and tools, e.g., embedded microcontrollers, GPUs; Windows, UNIX, Linux; git, Make, subversion, glibc, gcc, bash, etc

### HOT TOPICS

**Analytics Competency Center** Open Source Code **Data Governance** Data Scientist Innovation **Big Data** Hadoop Cloud IoT

Discovery **Data Prep** Machine Learning/Al Storytelling **Open Source Stack** Cybersecurity **Data Strategy Talent Management** Connected "X"



### **Embracing Big Data**



- Data is the new bacon ... but more bacon isn't always good for us
- Begin with the end in mind ... get your results into production
- Make analytics work at scale: automated, multi-user, secure
- Analytic skills: [CS] + [math] + [plain spoken] + [domain knowledge]  $\rightarrow$  \$xxx,xxx



## Questions / Ideas?

lonnie.miller@sas.com @AutosRCool

