2018 Top 10 Business Intelligence Trends
Don’t Fear Artificial Intelligence
1. Don’t Fear Artificial Intelligence
2. Liberal Arts Impact
3. Promise of NLP
4. Multi-Cloud Debate
5. Rise of the Chief Data Officer
6. Crowdsourced Governance
7. Data Insurance
8. Data Engineer Role
9. Location Internet of Things
10. Academics Investment
Don’t Fear Artificial Intelligence
Moderator & Speakers

Andy Cotgreave
Technical Evangelism Director
Moderator

Chris Fraley
Senior Research Scientist

Eric Brochu
Senior Software Engineer
Who we are

Andy Cotgreave

Andy is a visual analytics expert who has been with Tableau since 2011 in various roles ranging from product consultant to social content manager. He is now Tableau's technical evangelism director. Prior to Tableau, he was a data analyst at the University of Oxford.

As a technical evangelist, Andy helps people see and understand their data using Tableau's innovative products. He shares his passion for visual analysis and technology with his writing, (e.g. Computerworld, on tableau.com, and his own blog), speaking at industry conferences like SXSW and Tableau’s own events.
Who we are

Eric Brochu

- PhD in CS from University of British Columbia
  - Thesis in Machine Learning, worked on models for nonlinear optimization problems using human input
- Joined Tableau in March 2016
  - Working on Recommendations team, helping design and implement recommender systems for tables & joins and data sources.
Who we are

Chris Fraley

- PhD in CS from Stanford University
  - 25+ years experience in statistical computing
- Joined Tableau in November 2015
  - Working on Analytics research and development unit
  - Contributed to Tableau’s clustering algorithm
Agenda

1. Don’t Fear Artificial Intelligence
2. Machine Learning Definitions
3. Advantages & Limitations with Machine Learning
4. Machine Learning at Tableau
5. Open Discussion with Panel
Machine Learning Introduction
AI & Machine Learning To Boom

IDC forecasts revenues from AI and machine learning systems to total $46 billion by 2020.

In 2020, AI will become a positive net job motivator, creating 2.3 million jobs, Gartner reports.
“The development of full artificial intelligence could spell the end of the human race.”

- Stephen Hawking
  (BBC Interview, 2014)
“The development of full artificial intelligence could spell the end of the human race.”

- Stephen Hawking (BBC Interview, 2014)
“I don’t work on not turning AI evil today for the same reason I don't worry about the problem of overpopulation on the planet Mars.”

Andrew Ng,
VP & Chief Scientist, Baidu
AlphaGo Zero
Starting from scratch
AlphaGo Zero
Starting from scratch

Deep Thinking
Where Machine Intelligence Ends
And Human Creativity Begins

Garry Kasparov

‘Read this and embrace the future’ WALTER ISAACSON

‘Optimistic, wise and compelling’ CHARLES Duhigg
Robots will take a third of British jobs by 2030, report says

The Starship delivery robot has already started transporting food orders

The robots are coming – but will they really take all our jobs?

Artificial intelligence is going to transform the world, and our lives. But are we heading for a brave new world, or a science fiction horror-show?

David Barnett | Thursday 30 November 2017 06:15 GMT | 028 comments

Last week, Chancellor Philip Hammond announced in the Autumn Budget a £4bn package of investment into tech
Up to 30% of existing UK jobs could be impacted by automation by early 2030s, but this should be offset by job gains elsewhere in economy

Mar 24, 2017
Machine Learning Definitions
What is Machine Learning?
What is Machine Learning?

Learn from experience

Follow instructions
How does it differ from AI? Data science? Deep learning?
• Machine learning
  - Methods which are trained on past data to produce future output
• Machine learning
  – Methods which are trained on past data to produce future output
• AI
  – Mimicking human understanding and behavior
  – No requirement to use ML
  – ML happens to work the best
• Machine learning
  – Methods which are trained on past data to produce future output

• Deep learning
  – Specific collection of Neural Net based ML methods
  – Pretty amazing results on certain problems
• Machine learning
  – Methods which are trained on past data to produce future output
• Data science
  – Full pipeline to solve problems with data
  – Often uses ML, focuses less on automation
• Statistics
  – Same / similar techniques
  – Usually different goals
Advantages / Limitations of ML
Advantages of Machine Learning

#1: Don’t have to develop math/stat algorithms from scratch
Advantages of Machine Learning

#2: Speed up decision making through automation
Advantages of Machine Learning

#3: Learn automatically through examples (e.g. classification)
Advantages of Machine Learning

#4: Discover patterns, like clustering/segmentation or outlier/anomaly detection
Advantages of Machine Learning

#4: Discover patterns, like clustering/segmentation or outlier/anomaly detection
Advantages of Machine Learning

#5: Predict outcomes

1. Define Business Objectives
2. Identify Churn Parameters
3. Churn Prediction Model Building and Testing
4. Predict and Score Customers Likely to Churn

Risk of Churn:
- Low Risk
- Medium Risk
- High Risk
Advantages of Machine Learning

#6: Identify which inputs most affect outcomes

Customer churn – why do customers change operators?

The top 3 reasons why subscribers change providers:

- They want a new handset
- They believe they pay too much for calls
- Providers do not offer additional loyalty benefits

Source: Research conducted by Analysys Mason and Buongiorno
Limitations of Machine Learning

#1: Doesn't eliminate the importance of quantity and quality of data collected
Limitations of Machine Learning

#2: Need for validation and uncertainty quantification
#3: Decisions to be made when using machine learning methods:

**Data transformation**
- Model selection
- Variable selection
- Parameter settings
- Handling missing data
- Handling outliers/anomalies

Limitations of Machine Learning
Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

- Data transformation
- **Model selection**
- Variable selection
- Parameter settings
- Handling missing data
- Handling outliers/anomalies
Limitations of Machine Learning

#3: Decisions to be made when using machine learning methods:

- Data transformation
- Model selection
- **Variable selection**
- Parameter settings
- Handling missing data
- Handling outliers/anomalies

![Diagram of machine learning process](image-url)
#3: Decisions to be made when using machine learning methods:

- Data transformation
- Model selection
- Variable selection
- Parameter settings
- Handling missing data
- Handling outliers/anomalies

Limitations of Machine Learning
#3: Decisions to be made when using machine learning methods:

- Data transformation
- Model selection
- Variable selection
- Parameter settings
- Handling missing data
- Handling outliers/anomalies
Limitations of Machine Learning

#4: Machine Learning models can be hard to interpret
Machine Learning at Tableau
Smart @ Tableau

Human-Centered
- Augmentative.
- Empowering.
- Focus on things machines are good at and humans are not.
- Don’t interrupt me!

Contextual
- What is the task?
- Knowledge elicitation.
- What to infer and how?

Adaptive
- “Learning”
- More data = things can improve.
- The more we can learn, the better for everyone.

[T]he computer is the most remarkable tool that we’ve ever come up with. It’s the equivalent of a bicycle for our minds.

~ Steve Jobs, Memory & Imagination (1990)
Ways to be Smart

Heuristics/rules

UX

Data Science

Machine Learning
Machine Learning: a very very short intro

Data (+ Labels) -> training -> Model -> generalization -> Predictions

Cat, Not-cat

Cat, Not-cat
Machine Learning Key Points

Data (+ Labels)  \(\rightarrow\)  training  \(\rightarrow\)  Model  \(\rightarrow\)  generalization  \(\rightarrow\)  Predictions

adaptation

Data-Driven
Generalization
Adaptive
Being data-driven, generalizable and adaptive doesn’t just let us do things better, it enables us to do things that were previously impossible (or at least impractical).
Machine Learning Key Points

BUT, it’s not going to solve all our problems
- Requires formally-specified problems
- Requires “good” data, and plenty of it

How to get from data + questions to features?
Smart/ML = big changes coming to BI

Maximize Daily Quantity in the Customer-Data-QA46 Dataset

DailyQuantity by Store when Promotion is Display

San Francisco when Display

331 Store

1,530 Average

594,430 Total

Standard Deviation: 1,004

Difference from Mean: 1,195

Here are some cases where DailyQuantity was better than average:

- San Francisco is $25,35 above average
- Boston is $463.30 above average
- Little Rock is $280.00 above average

Here are some cases where profit was worse than average:

- New York is $359,91 below average

What else is interesting about these fields?

- TechSupport and OnlineSecurity are strongly associated.
- Average MonthlyCharges differs across InternetService.
- Tenure and Technology are strongly associated.
Tableau Recommendations
Recommendations

• What did we learn from working on recommendations?

• How did this help us think about ML? What decisions did we make?
Recommendations

OFFLINE

- Training Process
- Data
  - Users
  - Worksheet views
  - Table joins
  - Data sources

ONLINE

- Context
- Decision Process
  - Active user
  - Current connection
  - Tables and joins
  - Recommended Items
Abstract Machine Learning system

Data (+ Labels) → training → Model → generalization → Predictions

Training Process → Model → Decision Process

Context

Data

Tableau Recommendations system
Abstract Machine Learning system

Data (+ Labels) → training → Model → generalization → Predictions

- Data
- Training Process
- Model
- Decision Process
- Context
- Recommended Items

Tableau Recommendations system
Table & Join Recommendations

- Surface popular tables and joins
- Training data = table join data
- Model = associations
- Prediction = popularity
Data Source Recommendations

• Model = “People like you tend to like this.”

• Training data = user-workbook interactions

• Prediction = how likely user is to interact with item
6. Table join recommendations from foreign keys and database views

1. Common data prep problem: string deduplication
   - When data is typed in manually, especially by multiple individuals, it is bound to have inconsistencies in spelling and/or punctuation. This problem can waste many hours trying to coalesce all the different versions to a single one.
   - The approach that provided reasonable results in a timely manner was the token collision with double metaphone.
   - In a Nutshell: every string is tokenized by the double metaphone algorithm, which makes similar sounding strings

2. Automatically Finding Joinable Datasources
   - An experiment to determine if, through data analysis, existing Azure Server Datasources can be automatically joined with the users current datasources.
   - Techniques using Bloom Filters were shown to be very promising.
   - Discovered with Anurakha to determine applicability and the approach will be used in Maestro.

3. Automatic Statistician
   - Automatic Statistician is a tool that uses Machine Learning techniques to build an artificial intelligence system for data science.
   - It analyzes data points in a semi-automatization of functions that interested attributes of the periodically noize change so it communicates it to an analyst.
   - There is some existing code from Machine Learning Group.

4. Content Model Graph Search
   - Storing the full TWiki/TDWS content in a graph repository.
   - Each entry is an element (datasource, connection, field, worksheet).
   - Temporal connected to NoSQL.
   - Direct relationships between published or shared artifacts.
     - Example: A direct reference from workbooks to their published data.
     - Example: Connections to the same DB are shared.
   - Content searches in this model can specify exactly where they want to search.
   - Example: Find all cells where Excel vs find in title.
   - Content searches in this model can depict connections.
     - Example: Ensure that all terms are descendants of a single Workbook.

5. Smart Pills
   - Smart Pills highlights fields related to those that you are working with so that you can get straight to the most insightful visualizations of your data. As you drag fields to the rows and columns shelves, relevant fields that you have not yet considered are highlighted so that you know there is an interesting relationship to explore.
   - In real-time, an association rule mining algorithm that is discovering statistically significant associations between the fields you are considering and those you haven’t yet looked at.
   - The workflow has similarities to Recommended Table & Loops, but for fields instead of tables.

---

KDA: Key drivers/influencers analysis (Live demo)
Open Discussion with Panel
Discussion #1:

When it comes to machine learning for my organization, where do we start? How do I know it is the right investment? What should my team be focused on?
Discussion #2:

How do you see natural language processing (NLP) playing into machine learning?
Discussion #3:
What gets you most excited when we can combine the power of machine learning and data analytics?
Additional Resources
BI Trends Additional Resources

2018 BI Trends – Facebook Live w/ Chief Product Officer Francois Ajenstat
http://tabsoft.co/2D653LF

BI Trends “11th Trend” – Facebook Live
http://tabsoft.co/2mlkBzV
Thank you