# SAS® GLOBAL FORUM 2019

**USERS** PROGRAM

APRIL 28 – MAY 1 | DALLAS, TX

# Michael P Dessauer

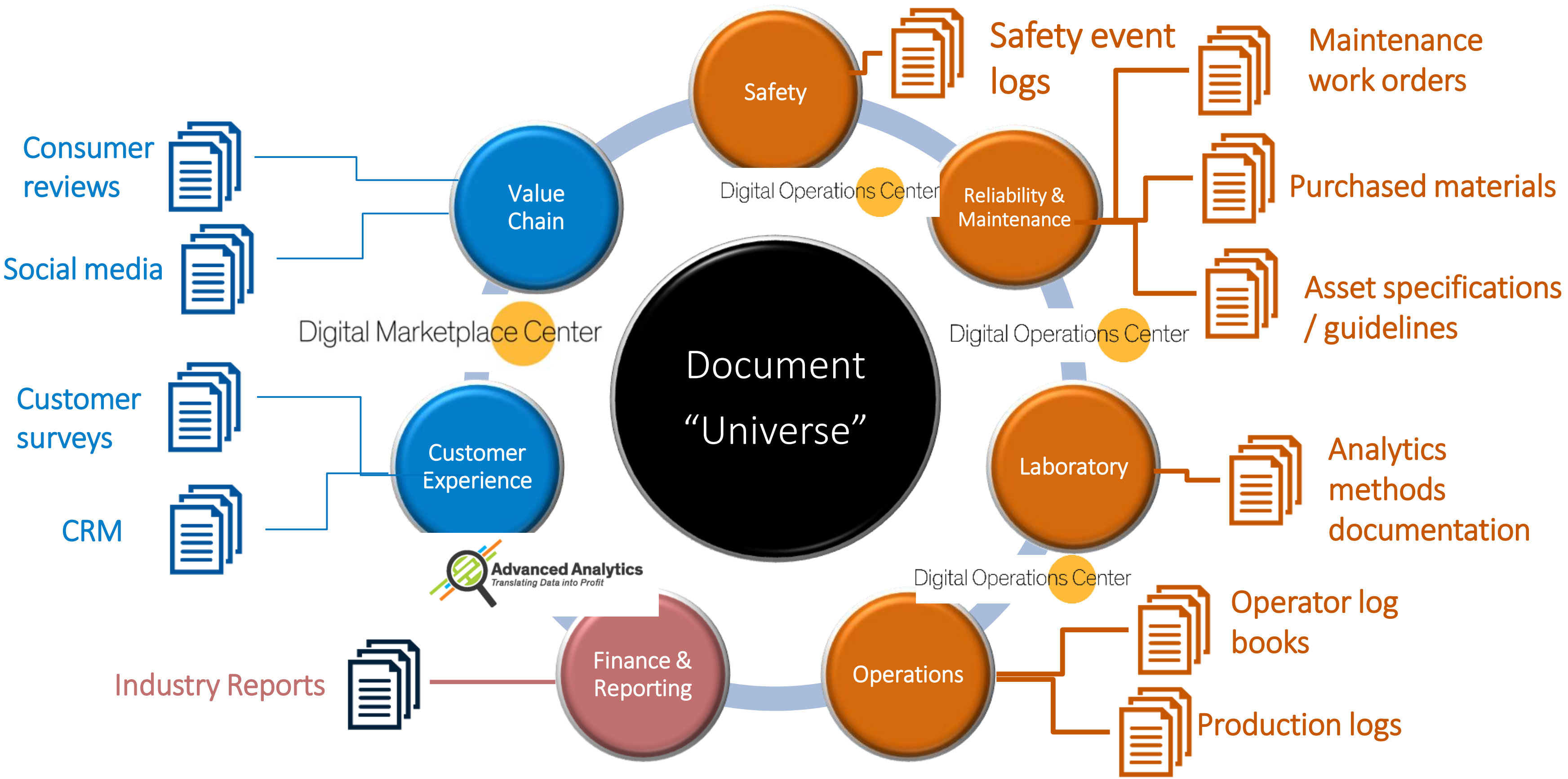
Michael Dessauer is the Analytics Lead Architect Specialist in Dow’s Digital Operations Center. Michael is currently focusing on manufacturing analytics areas such as predictive maintenance, GIS analytics, and text categorization for environmental health & safety, asset reliability, and maintenance organizations. Michael has a PhD in Computational Analysis and Modeling from Louisiana Tech

# Delivering Value Through Text Analytics in the Materials Manufacturing Industry

The Dow Chemical Company

# Materials Manufacturing Industry Document Universe (Abridged)

#SASGF



# Value Drivers for Materials Manufacturing

## What value outcomes can be driven from text-analytics insight?

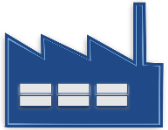
# Manufacturing - Let's introduce some questions…

## Safety

- How can I get immediate safety trends and alerts that are based off of free-text safety events?

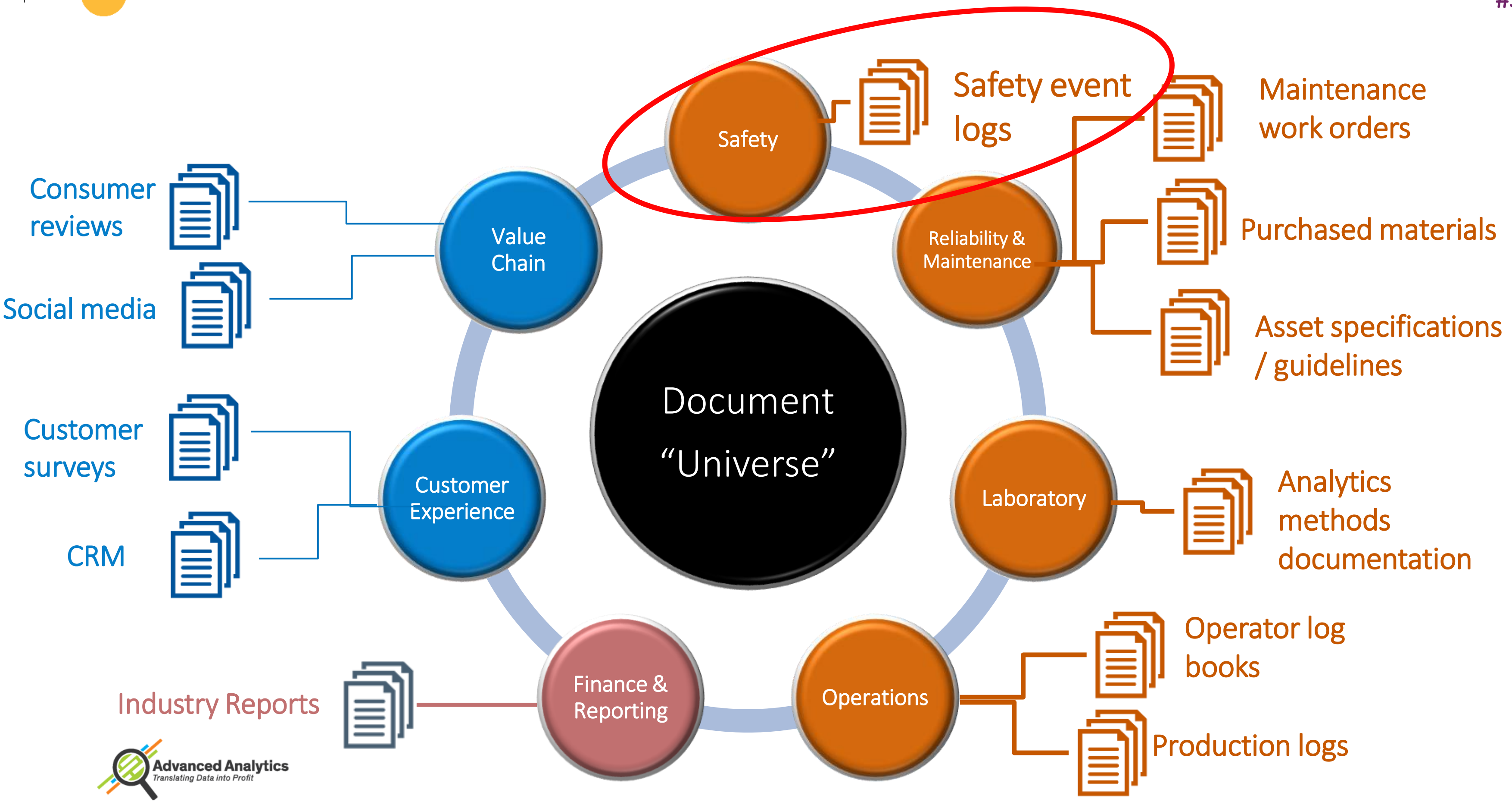
## Maintenance

- Can we use repair records and purchased materials to recommend future maintenance actions and predict failures?



## Production

- Can I use maintenance records and operator logs to understand equipment impact on production events?


Document
"Universe"
Safety
Safety event logs
Reliability & Maintenance
Maintenance work orders
Purchased materials
Asset specifications / guidelines
Laboratory
Analytics methods documentation
Operations
Operator log books
Production logs
Finance & Reporting
Industry Reports
Customer Experience
Customer surveys
CRM
Value Chain
Consumer reviews
Social media

# Case #1 – Automate Text Categorization for Safety

## Current Situation:

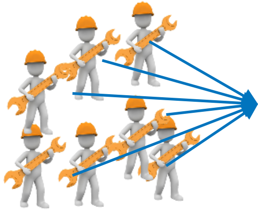
- Our site personnel must input a witnessed occurrence of unsafe work practices or conditions. These include **free text description** of the event.

## Challenge and Value

- How to understand trends that our safety organization can act upon in a timely manner? Need to **categorize and aggregate** the these free-text event records.

# Case #1 – Automate Text Categorization for Safety

## Historical Safety Event Trend Analysis Process



Site personnel inputs event text

Centralized repository

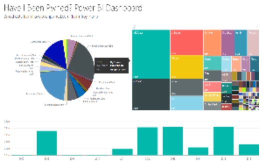
Quarterly Extract


Safety Organization Review

Manually Categorize Event

| Event ID | Category |
| --- | --- |
| 000000001 | A |
| 000000002 | B |
| 000000003 | A |
| 000000004 | C |
| 000000005 | D |

Generate Trend Dashboard



Time consuming & offline task

Can we automate this step????

# Case #1 – Automate Text Categorization for Safety

Can we automate this step????


Safety Organization Review

Yes!

## Automation for Categorizing Safety Events

Categorized Data

| Event ID | Text | Category |
|---|---|---|
| 000000001 | There was a safety issue | A |
| 000000002 | This was | B |
| 000000003 | Safety problem with helmet | A |

Train ML Model

New unlabeled text

| Event ID | Text |
|---|---|
| 000000001 | There was a safety issue |
| 000000002 | This was |
| 000000003 | Safety problem with helmet |

Categorized Documents

| Event ID | Category |
|---|---|
| 000000001 | A |
| 000000002 | B |
| 000000003 | A |

# Case #1 – Automate Text Categorization for Safety

Digital Operations Center

## ML Model Building

**Data conditioning**

| Text | Category |
|---|---|
| Worker witnessed another worker not tethered while on scaffold. | A |

Ensure text is converted to appropriate level of analysis (sentence, multi-sentence, phrase)

**Text Preprocessing**

| Text | Cat. |
|---|---|
| [Worker,(n)], [witness,(v)], [[not, tether(v)], …. | A |

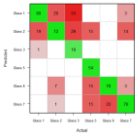
- Stop words
- Stemming
- Lemmatization
- Parts of Speech tagging

**Feature Transformation**

| Witness (n) | Witness (v) | Work (n) | Work (v) | Worker (n) | Cat. |
|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 1 | A |

- Transform processed text into Term-Document Matrix
- Word Embeddings other option (covered next)!!!

**ML Training**



- Library of multi-class classification models (SVM, NN, RandomForest)
- Train/Test split or K-fold cross validation

# Case #1 – Automate Text Categorization for Safety

## Conclusion of using ML to automate safety event categorization

- ~78% Accuracy for 2 sets of categorizations that have 16 different classes (not bad, but not satisfied quite yet)
- Still challenges using TDM (Term-Document matrices) because term **context is lost!**

# How do we use language "context" to achieve better model quality (and other neat functionality)?

| Traditional Method - Bag of Words Model | Word Embeddings |
|---|---|
| • Uses one hot encoding<br>• Each word in the vocabulary is represented by one bit position in a HUGE vector.<br>• For example, if we have a vocabulary of 10000 words, and "Hello" is the 4th word in the dictionary, it would be represented by: 0 0 0 1 0 0 . . . . . . . . 0 0 0 0<br>• Context information is not utilized | • Stores each word in as a point in space, where it is represented by a vector of fixed number of dimensions (generally 300)<br>• Unsupervised, built just by reading huge corpus<br>• For example, "Hello" might be represented as : [0.4, -0.11, 0.55, 0.3 . . . 0.1, 0.02]<br>• Dimensions are basically projections along different axes, more of a mathematical concept. |

# What actually are the word embeddings?

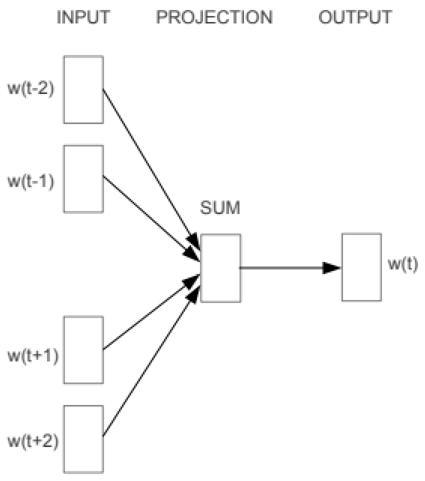
- The word embeddings are the “weights” from the hidden layer of a 2-layer neural net
- Actually <u>NOT</u> a deep neural net, but transforms text into a form deep neural nets can use
- To develop weights based on context, one typical method is the create a “continuous bag of words” (CBOW) model
- CBOW tries to predict a word based on its neighbors

**Example sentence:**

*the quick brown fox jumped over the lazy dog*

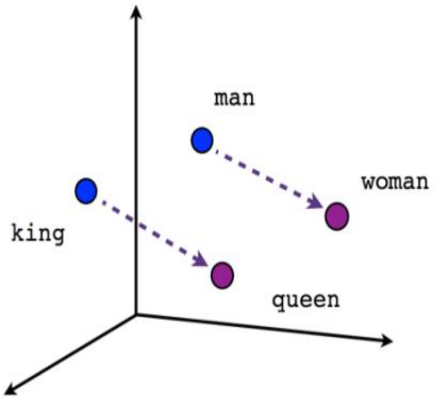
Using a windows size=2, we have a dataset:

*([the, brown, fox], quick), ([the, quick, fox, jumped], brown), ([quick, brown, jumped, over], fox), ...*

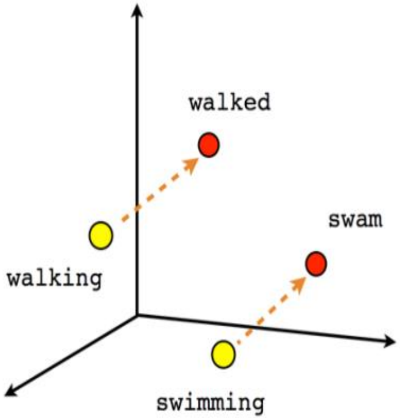



CBOW

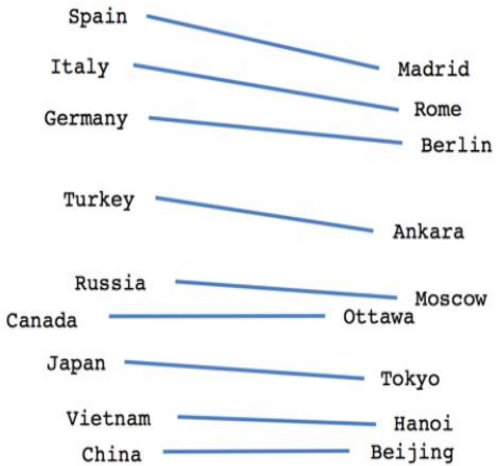
# Examples – Class Word2Vec Example




Male-Female



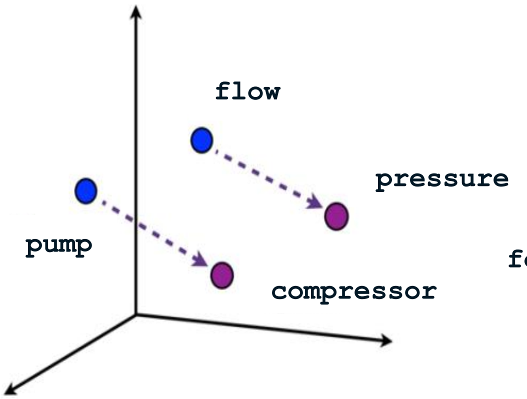

Verb tense



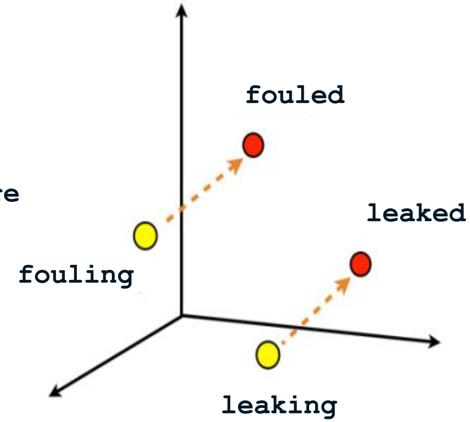

Country-Capital

vector[Queen] = vector[King] - vector[Man] + vector[Woman]

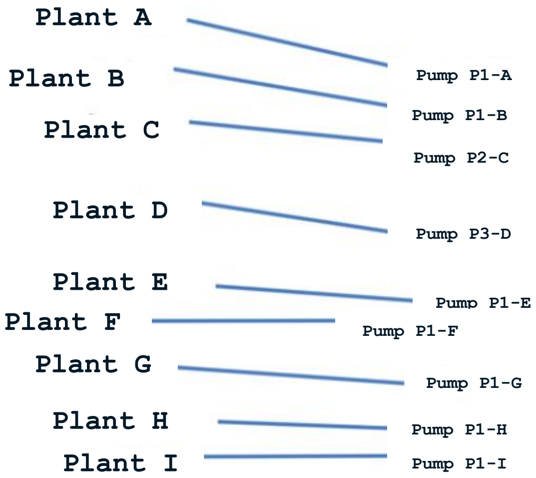
# Dow's version: DOW2VEC




Equipment - Process




Condition Verbs




Plant - Equipment

vector[Plant A] = vector[Plant B] - vector[Pump P-A1] + vector[Pump P-B1]

We can now leverage **context** within and between Dow document "worlds" to gain insights.

# Case #2 – Identifying Root Causes for Production Events

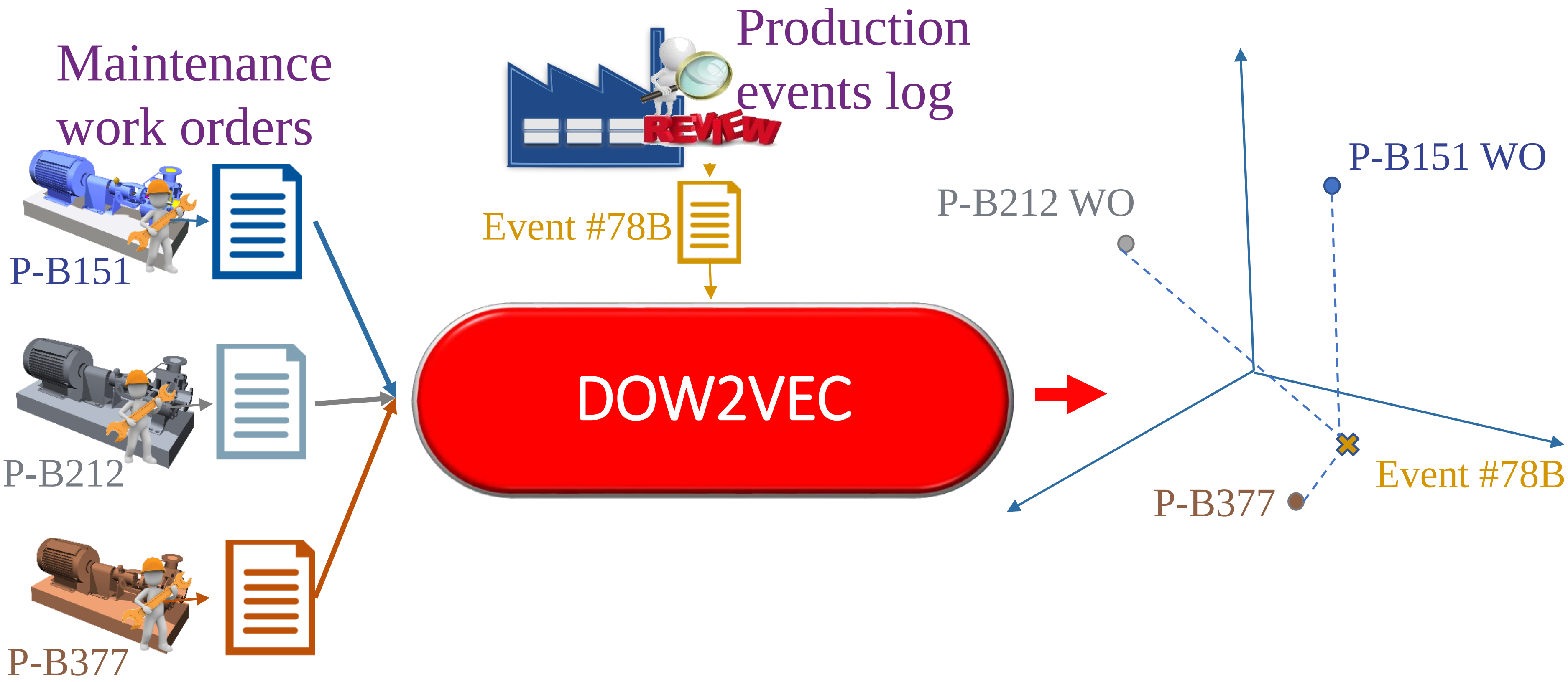
## Current Situation:

- Daily operations, maintenance, and production events are recorded using disparate systems. These systems use **free text descriptions** to describe event(s).

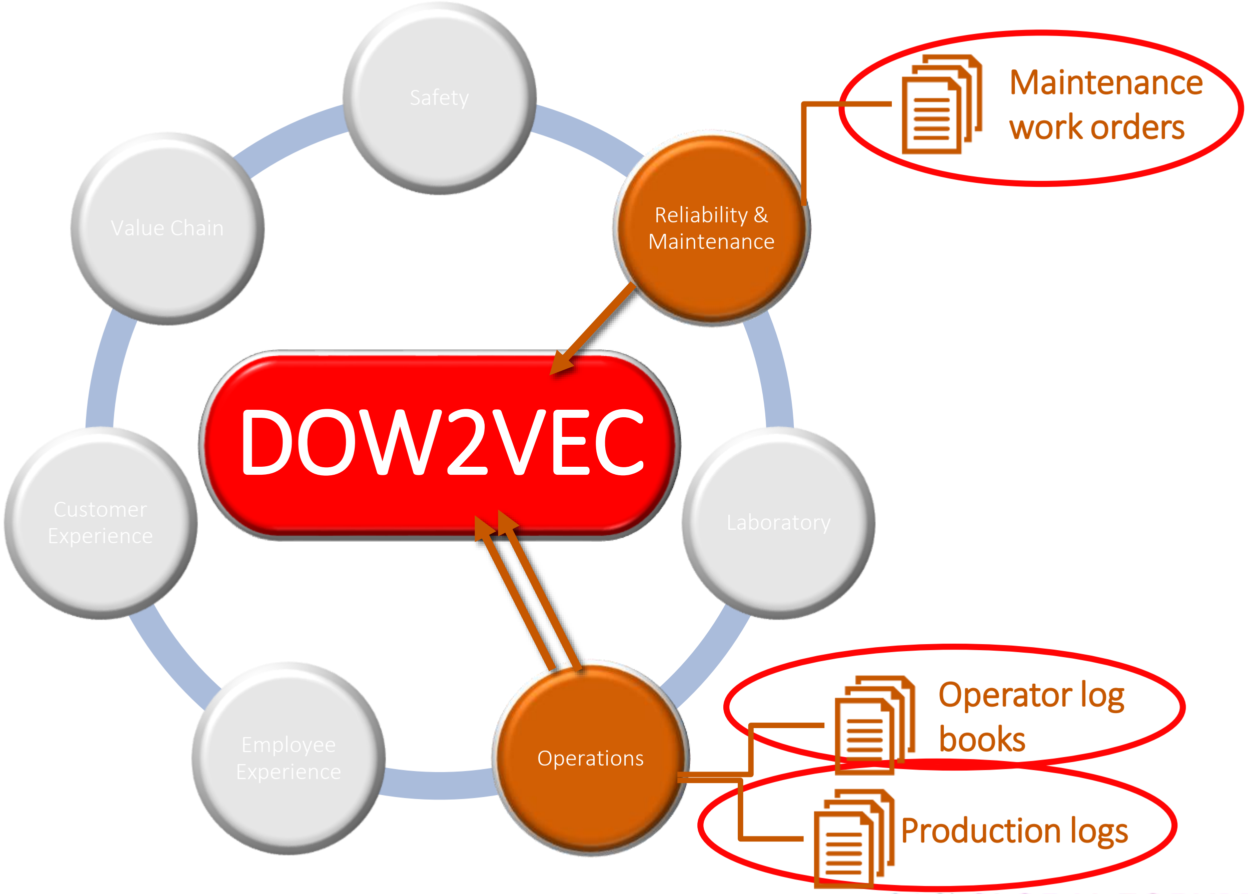
## Challenge and Value

- There is value in relating maintenance and operator logs to production events to better understand **root causes** and become **proactive** in mitigating future adverse events.

# Case #2 – Can we relate production events to equipment?



Develop n-dimensional space that can map logs to equipment repair data



# Can we use a vector space to train a model to predict equipment's impact on production events?

Operator log

Maintenance Records

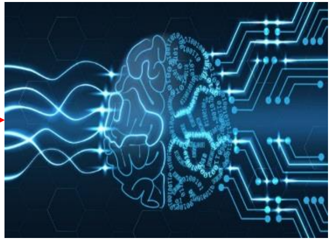
Production Logs

Transform text into "contextual" vector space

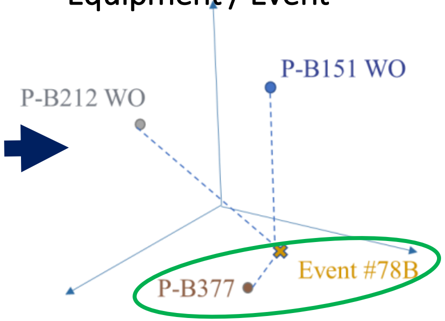
DOW2VEC

| d1 | d2 | d3 | d4 | d5 | … | d300 |
|---|---|---|---|---|---|---|
| .33 | .16 | .6 | .78 | .99 | … | .2 |
| .44 | .12 | .11 | .52 | .14 | … | .9 |
| .9 | .52 | .23 | .6 | .14 | … | .21 |
| .12 | .33 | .66 | .88 | | … | .4 |

Train model on maintenance records using Dow2Vec features

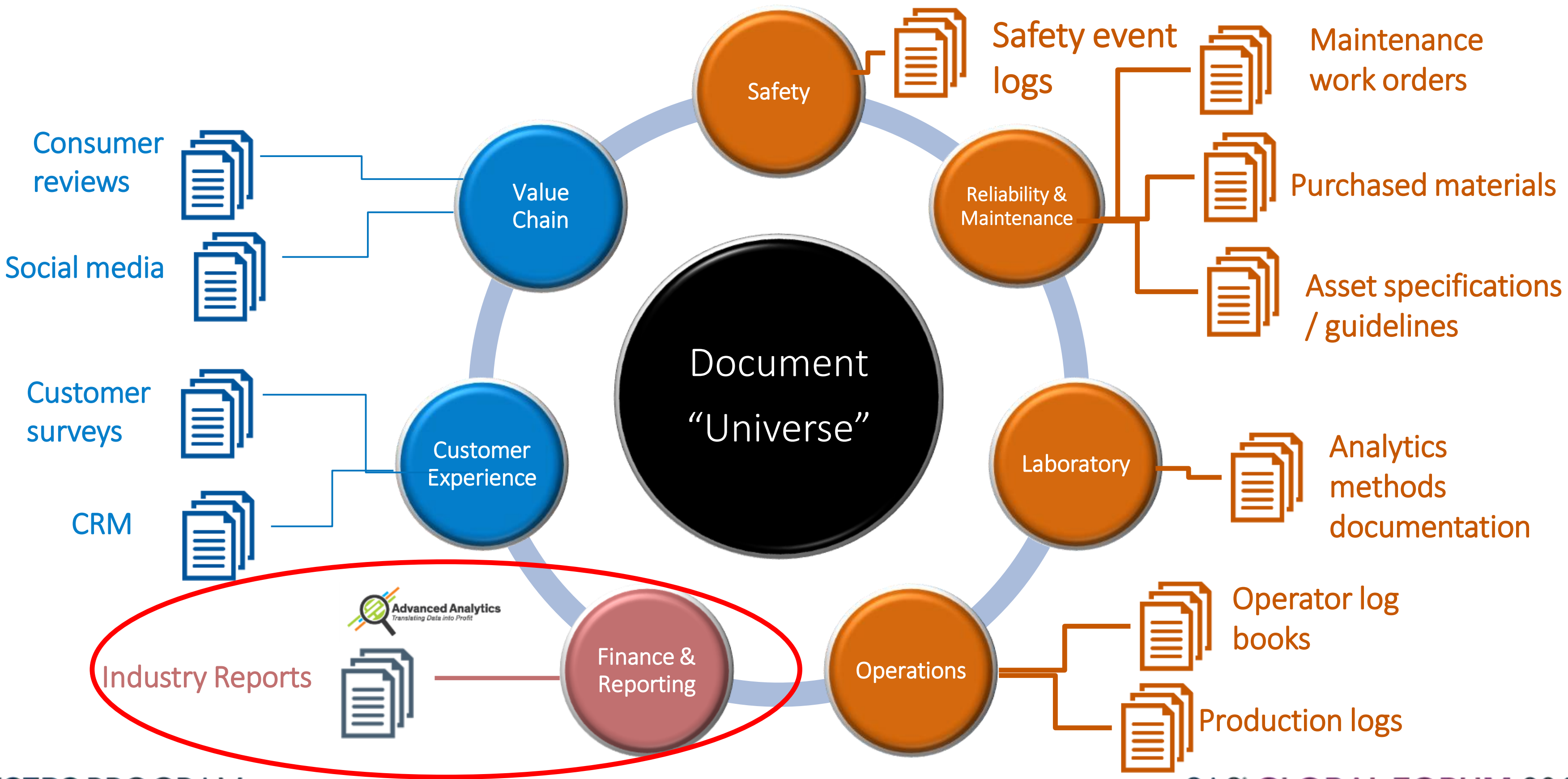


Equipment / Event



# Materials Manufacturing Industry Document Universe (Abridged)





# Case #3 – Industry Macro Trends Impact on Demand Forecasting

## Current Situation:

- Dow must schedule production and inventory based on customer needs through a demand planning process which can be an offline, qualitative process

## Challenge and Value

- How can macro economic trend information contained in industrial reports influence the demand planning process to improve production scheduling?

# Case #3 – Industry Macro Trends Impact on Demand Planning

## How to use external industry documents to improve demand planning?

- Provide a KPI that visualizes **topics sentiment trends** based on daily economic reports
  - Allows the news consumer to focus attention on topics with potentially changing sentiment
  - Gives a view of the past to put the current topic sentiment into context
- Can we use financial report topic sentiment to **improve models**:
  - Minimize Latency
  - Increase Quality
- Initial Datasource Size:
  - 1000s of Report PDFs
  - 100,000 sentences

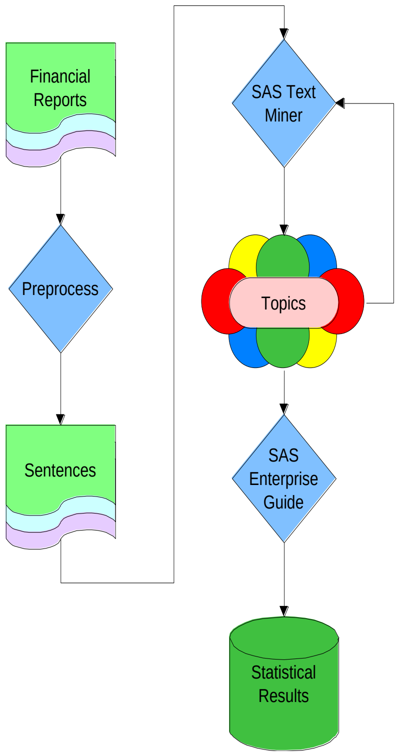
Advanced Analytics
Translating Data into Profit

# Text Mining Steps for Sentiment Scoring

Text Mining Steps for Sentiment Scoring

## Text Mining Steps

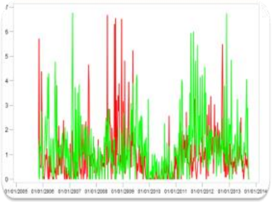
1. **Data import (Web Crawling to get links)**
   - Economic PDFs (Weekly/Daily)
   - Daily Updates PDF (Daily)
2. **Text Mining pre-processing (Python + Toolkits)**
   - Document Filtering
   - Sentence Parsing
3. **Text Mining Steps (SAS Text Miner)**
   - Term selections (Multi-word terms, Synonyms, PoS)
   - Topic Creation (Custom Topics, Refine w/ preprocess)
4. **Next Steps (SAS Text Miner / EG)**
   - Term / Topic Refinement
   - Sentiment Topics (Custom sentiment by Topic)



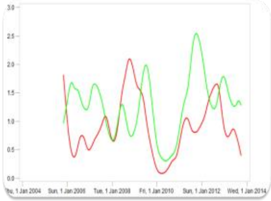
# Text Mining Steps for Sentiment Scoring

## Sentiment Scoring Post-Processing Steps

Daily Sentiment Scores By Geography/Topic

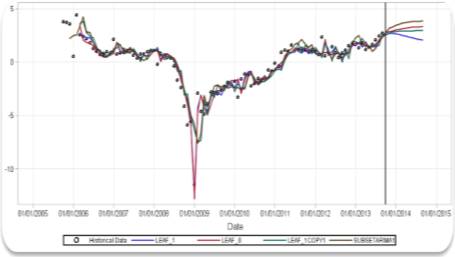


Aggregated Weekly Sentiment Trends



Analyze variable relationship with external indicators


Generate forecasts models using sentiment scores




**5. Aggregate Sentiment Scores into weekly time series**

- Take raw weights of sentiment and concept
- Binary geography topic

**6. Clean time series data**

- Linear spline interpolation
- Trend component
- Detect inflection points based on D1 sign changes

**7. Model external indicators (JMP, SAS FS)**

- Similarity / Co-integration
- Generate (SAS Forecast Studio) ARIMAX models using sentiment trends as inputs

# Case #3 – Industry Macro Trends Impact on Demand Planning



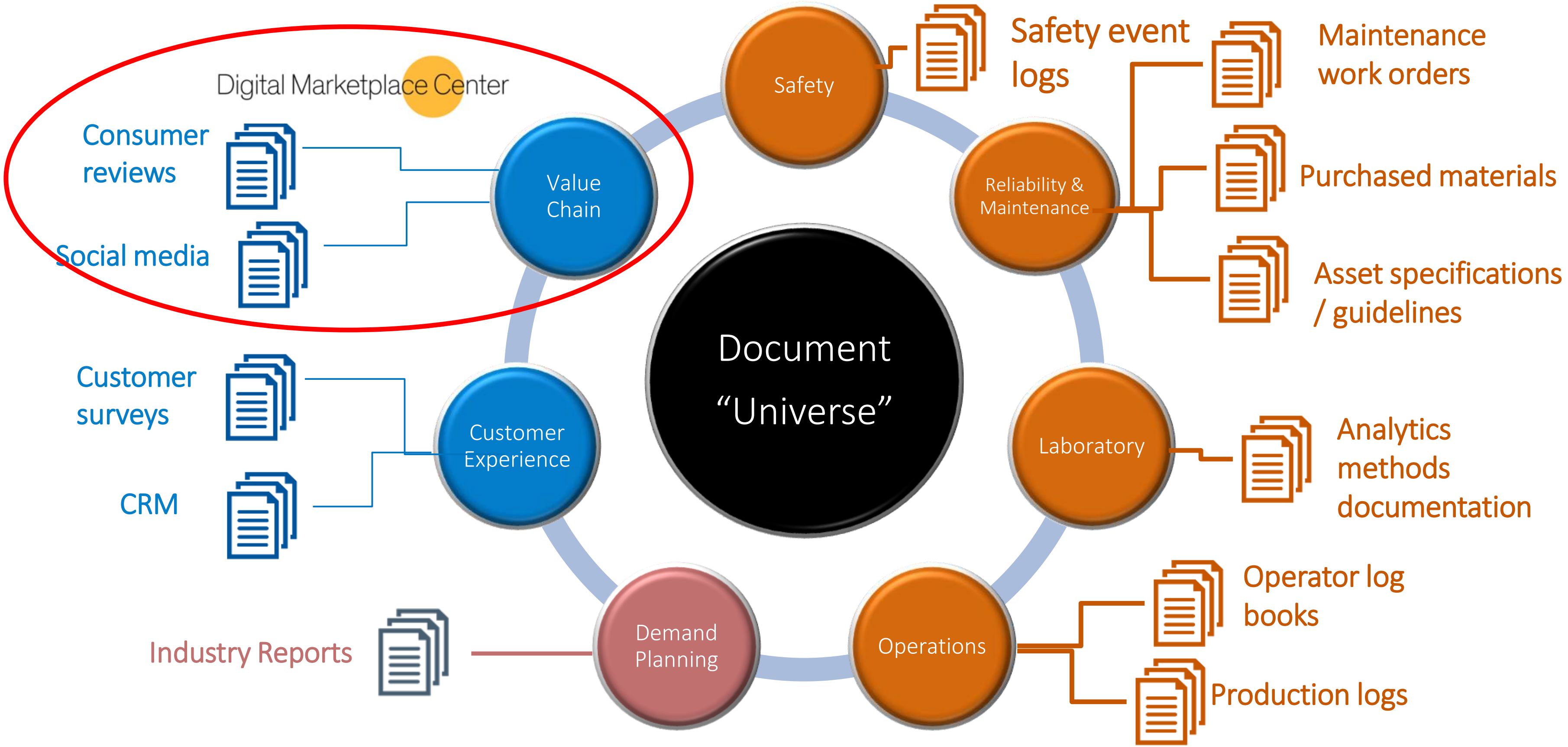
Advanced Analytics
*Translating Data into Profit*

## Conclusions

- Text-based sentiment inputs **can improve** demand forecasting models.
- Should be considered an **additional input** for sales forecasting.
  - Helpful when structured data is unavailable

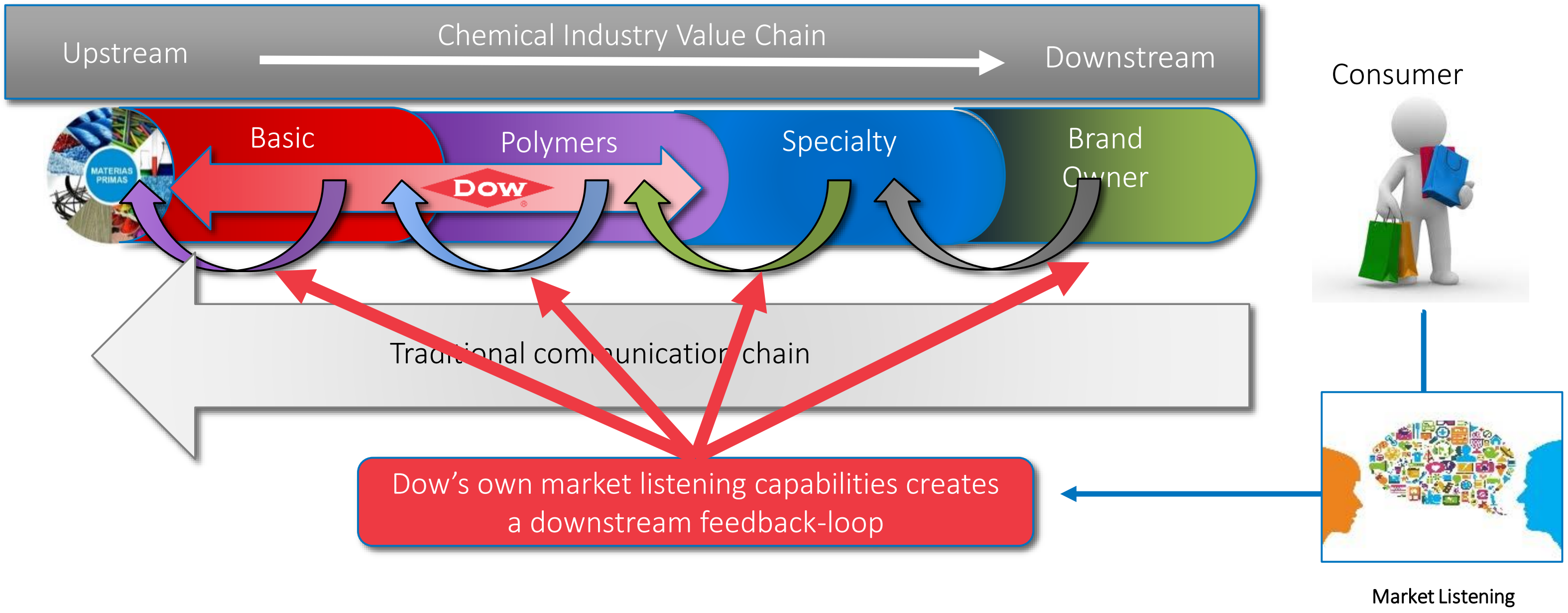
# Materials Manufacturing Industry Document Universe (Abridged)





Digital Marketplace Center

# Dow’s position in value chain




® The DOW Diamond Logo is a trademark of The Dow Chemical Company

# How can Dow generate value from Market Listening?

## Isolate Dow formulation's value through consumer sentiment

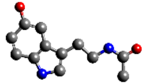

Search by
- Brand
- Product
- Attribute

Analyze 1000's of relevant documents

Digital Marketplace Center

**Product Attribute**

Brand 1 | Brand 2 | Brand 3

Compare products using Dow formulations verse other Brands

# Market Listening Project Process at Dow

Digital Marketplace Center

## Targeted Text Collection

Scrapy

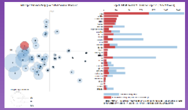
Internal & Vendor tools to search and collect text

## Entity, Topic, & Sentiment Categorization

Work with marketing experts to generate valuable categorizations

## Modeling & NLP

Natural Language Analyses with NLTK

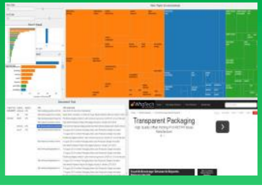


sas

scikits learn machine learning in Python

Modeling tools for categorization and/or extraction

## Insight Delivery




Deliver interactive dashboards with regular data updates

# Case Study – Strengthen our Value Proposition

- Brand-specific social media data sources
- Leveraged several tools for
  - Sentence parsing
  - Part-of-speech tagging
  - Topic identification
- Aspect-specific sentiment weighting
- Simple delivery

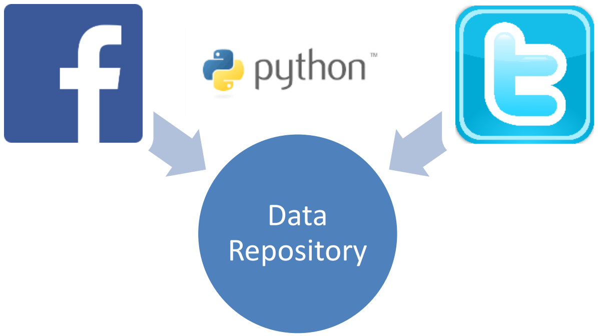
# Case Study – Relevant Text Collection

Digital Marketplace Center

Targeted Text Collection

Scrapy

Internal & Vendor tools to search and collect text



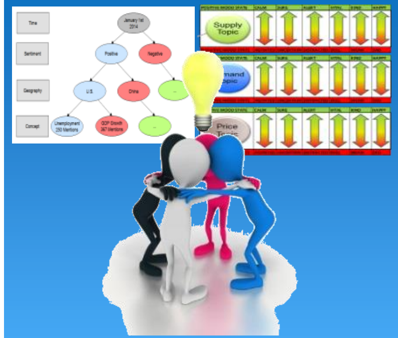

- Domain Expert targeted sources on Twitter and Facebook
- Data extraction using Python through public APIs
- Download to harmonized central repository

# Case Study – Value-driven Text Categorization

Digital Marketplace Center

## How to categorize into structured fields?

Entity, Topic, & Sentiment Categorization



Work with marketing experts to generate valuable categorizations

Brands?

Product Attributes?

How to define sentiment?

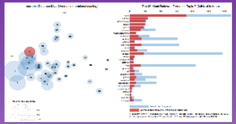
- Business Marketing and R&D-lead effort to determine value case
- Define entities, attributes, and sentiment to quantitatively assign value to consumer needs
- Feedback loop with modeling step to validate results
- Topic modeling methods: LDA or SVD (text topic in SAS Text Miner)

# Case Study – Modeling & Text NLP



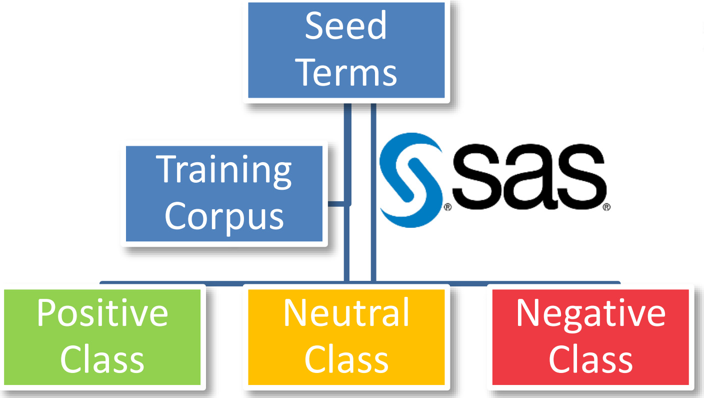
## How to develop models to categorize, summarize, identify topics?

Modeling & Text NLP

Modeling tools for categorization and/or extraction

Digital Marketplace Center




- Natural Language ToolKit (NLTK) leveraged for sentence level parsing
- SAS Text Miner used to develop sentiment classes
- Sentences weighted on sentiment content

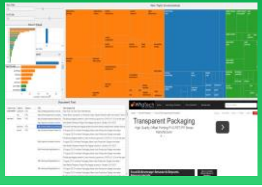
# Case Study – Deliver Insight

#SASGF

Digital Marketplace Center

## Insight Delivery – Make Actions!

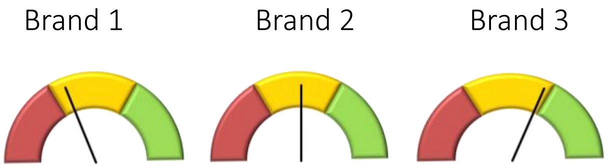
Insights Delivery




Deliver interactive dashboards with regular data updates

| post | sentence | Topic 1 Score | Topic 1 Sentiment | Topic 2 Score | Topci 2 Sentiment |
|---|---|---|---|---|---|
| post 1 | sentence 1 | 0.78 | -2.33 | 0.05 | 1.34 |
| post 1 | sentence 2 | 0.03 | 1.22 | 0.88 | -1.33 |
| post 1 | sentence 3 | 0.34 | 0.2 | 0.33 | 0 |

Aspect #1




- Simple, Ad Hoc, delivery method
- Each sentence in each post had an individual sentiment score for each of the aspects
- Sentiment was rolled up to the Brand level and reported based on relevance to the post

# Conclusions – Using text analytics to help solve problems in material manufacturing

## The Time is NOW

- With recent advancements in NLP-based modeling, Dow (and many other companies) can develop value-driven models and applications to enable our operations.
- for understanding how users can unlock existing document collections to generate insights faster.

## Power in Collaboration

Projects like Dow2Vec and market listening get more powerful as new use cases and document collections are included!

# Thank you!

Contact Information
mpdessauer@dow.com

# Reminder:

Complete your session survey in the conference mobile app.


#SASGF
SAS®
GLOBAL
FORUM
2019
APRIL 28 – MAY 1 | DALLAS, TX
Kay Bailey Hutchison Convention Center