DATA DRIVEN

TOYOTA Customer 360° on *Apache Spark™*

Brian Kursar, Sr Data Scientist
Toyota Motor Sales IT Research and Development
Final

TOYOTA Big Data History

R&D

Data Engineering

Infrastructure

Enterprise Architecture

SPARK SUMMIT 2014

現地現物

Genchi Genbutsu

“Go Look, Go See”

TOYOTA

- Compute
- Streaming
- Machine Learning

ACTIONABLE INSIGHTS

TOYOTA

Data Engineering

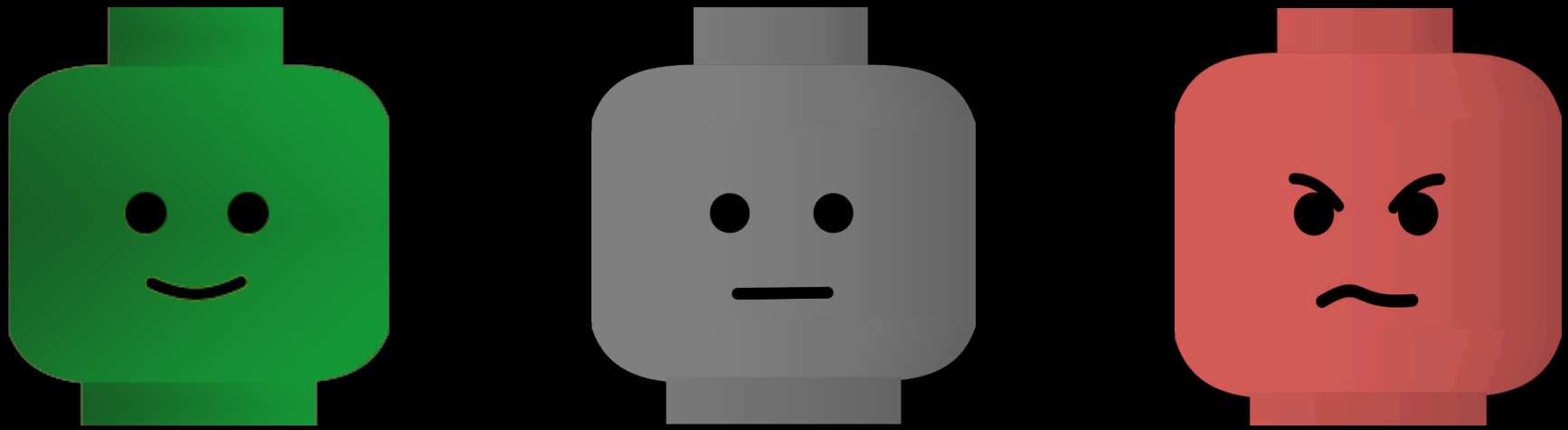
Customer Experience original Batch Job

160 hours (6.6 days)

Same job re-written using Apache Spark …

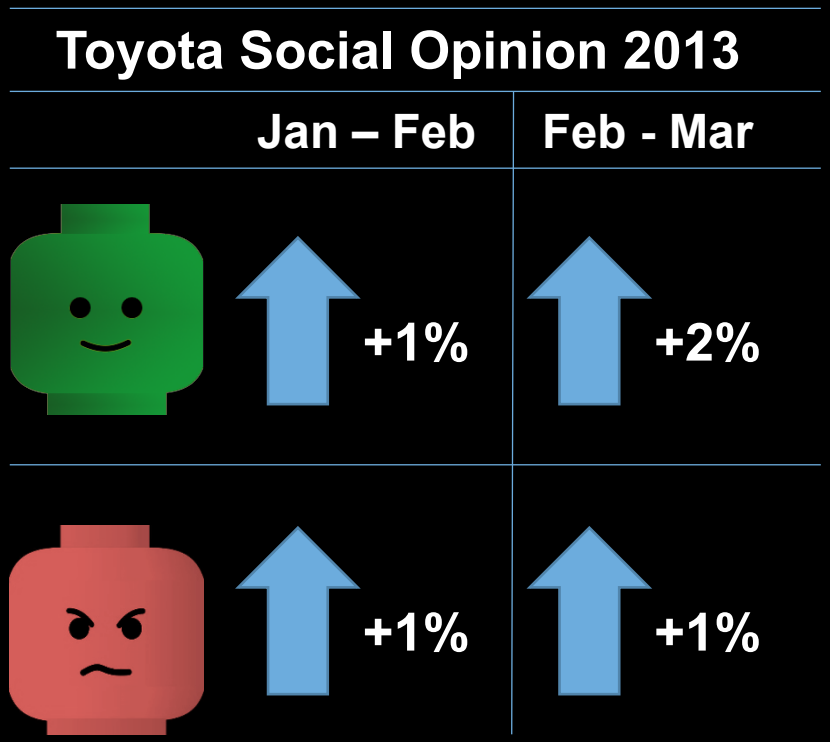
4 hours

Existing Tools



TOYOTA

Existing Tools



TOYOTA

Toyota Online Conversations by the Numbers

2014 Study

40%	**Retailers Selling Toyota Vehicles**
11%	**Opinions on Marketing Campaigns**
10%	**Feedback on Dealer Sales and Service Experiences**
9%	**Opinions on Product Styling and Features**
8%	**People In Market for a Toyota**
8%	**Incident Reports Involving a Toyota Vehicle**
7%	**Feedback on Product Quality**
5%	**Customers Advocating for the Brand**
2%	**Completely Irrelevant**

TOYOTA

Toyota Online Conversations by the Numbers

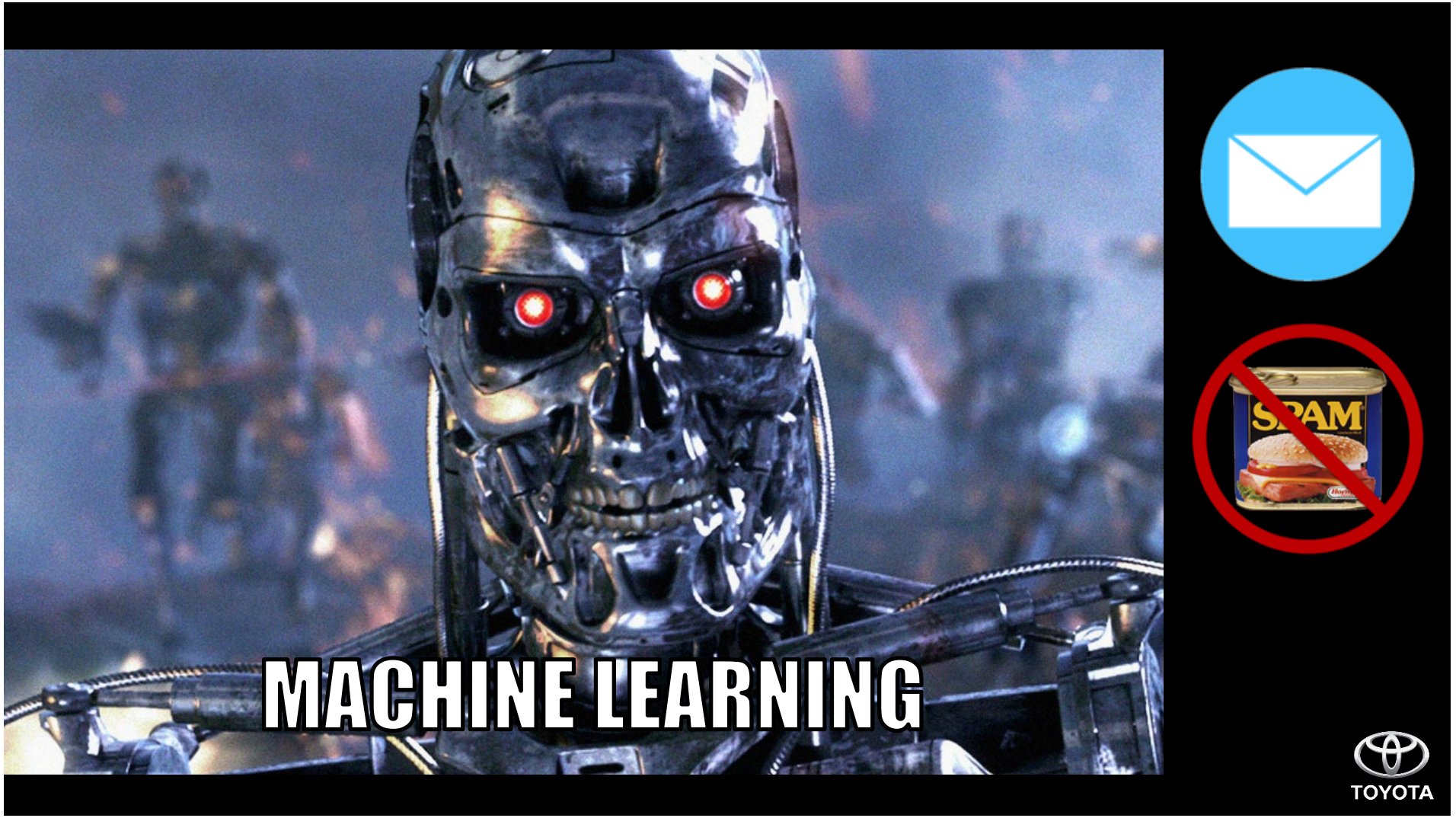
2014 Study

- 40% Retailers Selling Toyota Vehicles
- 11% Opinions on Marketing Campaigns
- 10% Feedback on Dealer Sales and Service Experiences
- 9% Opinions on Product Styling and Features
- 8% People In Market for a Toyota
- 8% Incident Reports Involving a Toyota Vehicle
- 7% Feedback on Product Quality
- 5% Customers Advocating for the Brand
- 2% Completely Irrelevant

50% Noise

TOYOTA


Spark
Categorize and Prioritize incoming Social Media interactions in Real-Time using Spark MLLib
Campaign Opinions
Customer Feedback
Product Feedback
Noise
TOYOTA


MACHINE LEARNING
SPAM
TOYOTA

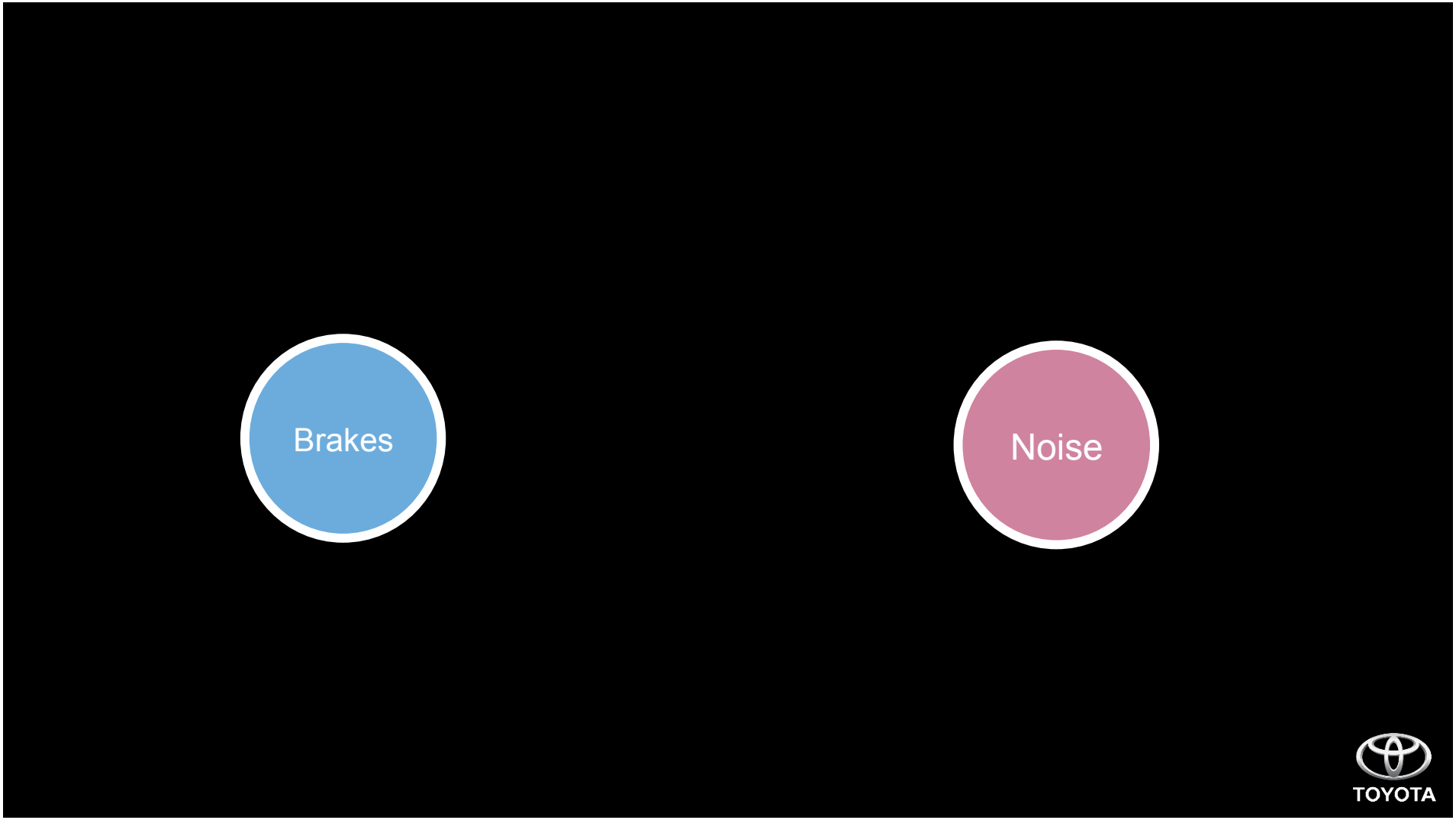
First Spark MLlib Experiment

Time-box project to 12 Weeks
Classify at min 80% accuracy

- Seat Cover Wrinkles/Cracking
- **Brake Noise**
- Shift Quality
- Oil Leaks
- HVAC Odor
- Dead Battery
- Rodent Wire Harness Damage
- Paint Chips

TOYOTA

Describe this issue…

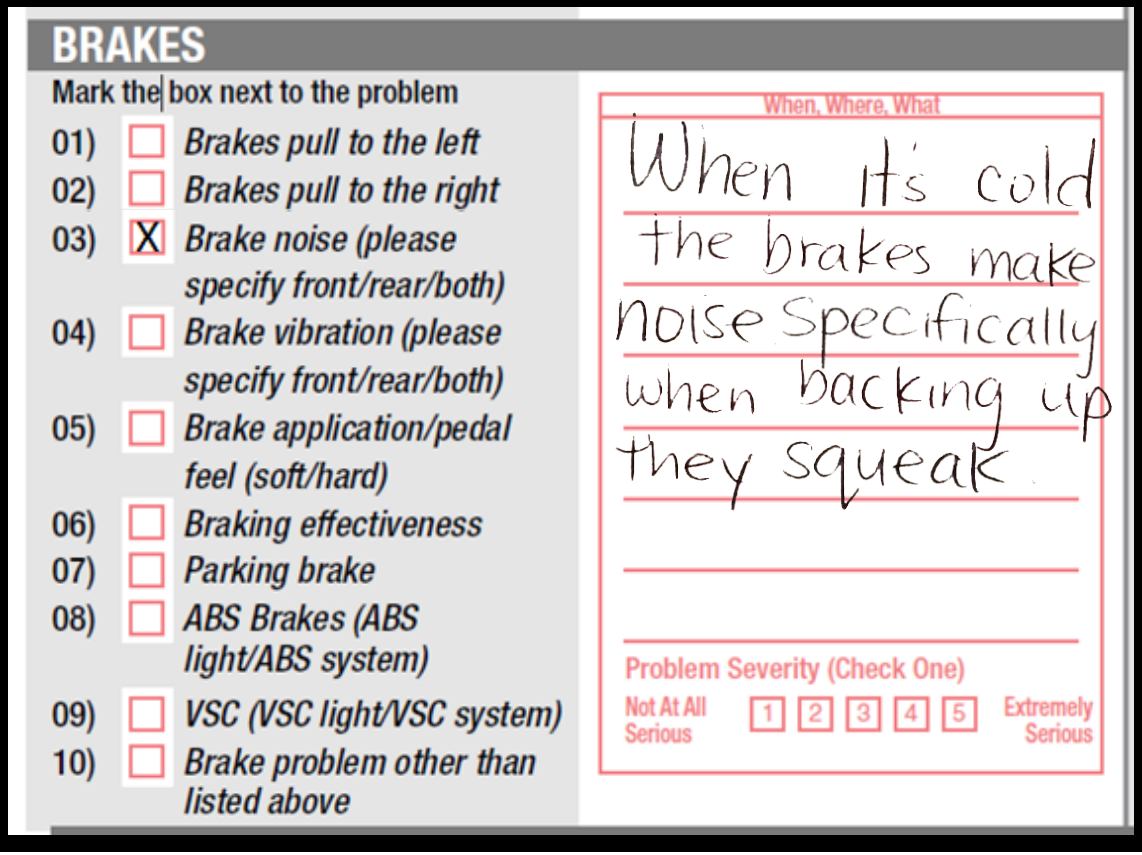

Brakes
Noise
TOYOTA

When I’m backing up in my 2012 Prius, it sounds like something hanging up or scraping as it rotates and only happens in the morning..

I hear a squeak coming from the back wheel of my Prius as I pull out from my driveway in the morning.

TOYOTA

Identify Training Data for Brake Noise Model



BRAKES

Mark the box next to the problem

01) ☐ *Brakes pull to the left*
02) ☐ *Brakes pull to the right*
03) ☒ *Brake noise (please specify front/rear/both)*
04) ☐ *Brake vibration (please specify front/rear/both)*
05) ☐ *Brake application/pedal feel (soft/hard)*
06) ☐ *Braking effectiveness*
07) ☐ *Parking brake*
08) ☐ *ABS Brakes (ABS light/ABS system)*
09) ☐ *VSC (VSC light/VSC system)*
10) ☐ *Brake problem other than listed above*

When, Where, What

When it's cold the brakes make noise Specifically when backing up they squeak

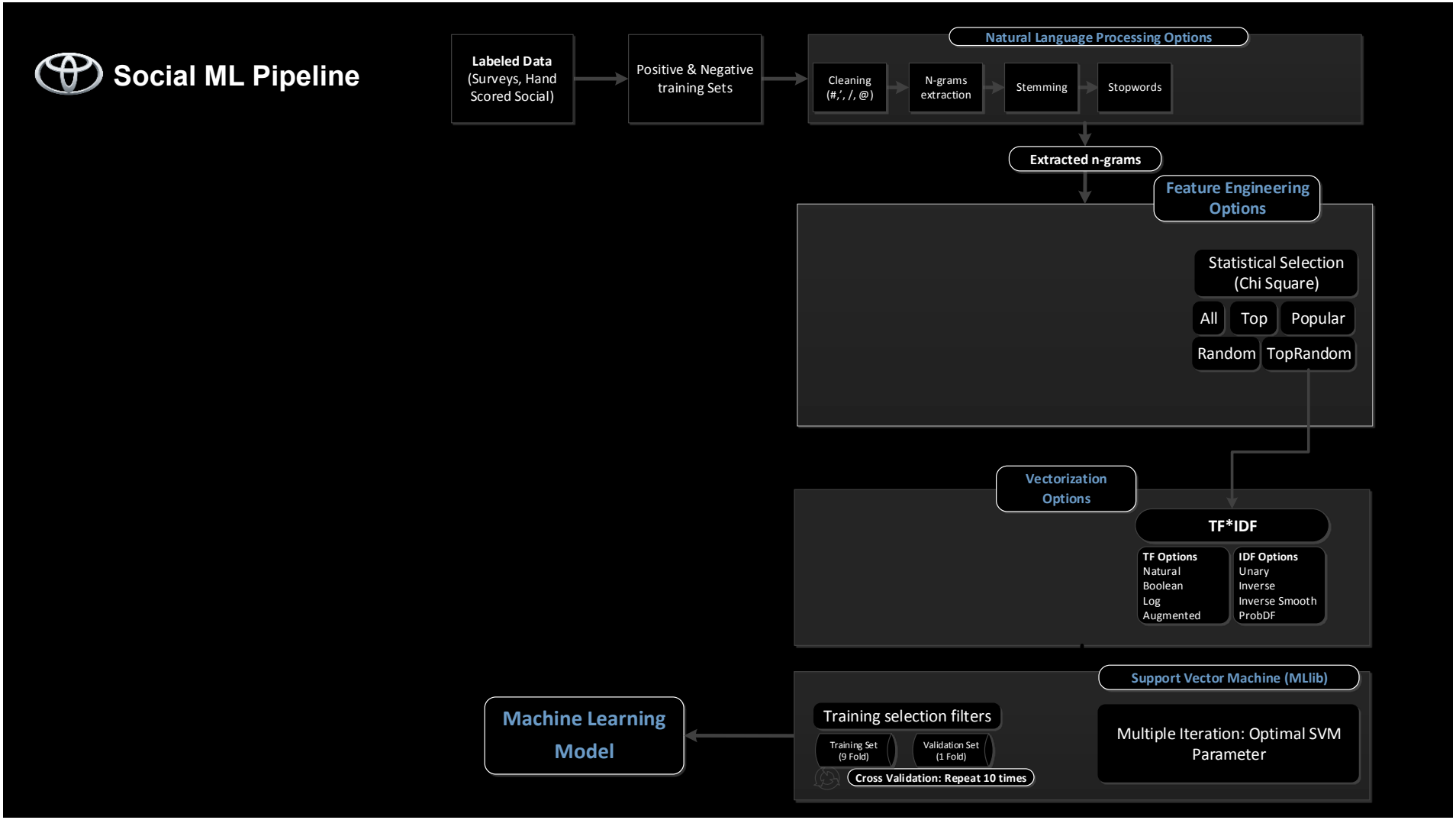
Problem Severity (Check One)

Not At All Serious 1 2 3 4 5 Extremely Serious

Extract Category (Label) matching Model Objective

Extract features from the Verbatim Text

TOYOTA


Social ML Pipeline
Labeled Data
(Surveys, Hand Scored Social)
Positive & Negative training Sets
Natural Language Processing Options
Cleaning (#,', /, @)
N-grams extraction
Stemming
Stopwords
Extracted n-grams
Feature Engineering Options
Statistical Selection (Chi Square)
All
Top
Popular
Random
TopRandom
Vectorization Options
TF*IDF
TF Options
Natural
Boolean
Log
Augmented
IDF Options
Unary
Inverse
Inverse Smooth
ProbDF
Support Vector Machine (MLlib)
Training selection filters
Training Set (9 Fold)
Validation Set (1 Fold)
Cross Validation: Repeat 10 times
Multiple Iteration: Optimal SVM Parameter
Machine Learning Model

Extract Text Features

```
object ChiSquareFeatureSelector {

  def selectFeatures(documents: RDD[LabeledNgramInfos], chiSquareCriticalValue: Double, rareTermThreshold: Int = 3): ChiSquareFeatureSelector = {

    val corpusSize = documents.count

    val classCountPerCorpus = countDocumentsByClassification(documents)

    val featuresRDD = documents
      .flatMap{ doc =>
        doc.features.map { ngram => (NgramClassification(ngram.term, doc.label), 1) }
      }
      .reduceByKey(_ + _)
      .map( x => (x._1.ngram, ClassificationCount(classificationKey=x._1.classificationKey, count=x._2 ) ) )
      .groupByKey()
      .flatMap { case(ngram, classificationCounts) =>
        calculateChiSquare(ngram, classificationCounts, classCountPerCorpus, corpusSize)
      }
      .filter(includeThisFeature(_, chiSquareCriticalValue, rareTermThreshold))
      .cache()

    new ChiSquareFeatureSelector(featuresRDD)
  }
```

TOYOTA

Train Predictive Model

```scala
object KFoldsTrainingModel {

  def train(
    labelName: String,
    kFolds: Int = 10,
    numSvmIterations: Integer = 10,
    svmGamma: Double = 0.5,
    svmRegParam: Double = 1.0,
    kernelType: String = "linear",
    negDocVectors: RDD[LabeledPoint],
    posDocVectors: RDD[LabeledPoint])(implicit sc: SparkContext) : KFoldsTrainingModel = {

    val votesRequired = math.ceil((kFolds.toDouble - 1) / 2)
    val foldsArray = Array.fill(kFolds)(1.0 / kFolds)

    val splits0 = negDocVectors
      .randomSplit(foldsArray, seed = 11L)

    val splits1 = posDocVectors
      .randomSplit(foldsArray, seed = 11L)

    val mlModels = kernelType match {
      case "linear" => trainKfoldsLinear(splits0, splits1, kFolds, numSvmIterations, svmRegParam)

    }

    // run predictions for each of the k-folds models
    val testFold = kFolds - 1
    val testData: RDD[LabeledPoint] = splits0(testFold).union( splits1(testFold) ).cache()
    val testModels = mlModels.take(kFolds - 1)

    kernelType match {
      case "linear" => {
        val predictParams = Seq(0.0)
        for( predictParam <- predictParams ) {
          predictAll(labelName, testModels, testData, votesRequired, predictParam.asInstanceOf[Double])
        }
      }

      case _ => predictAll(labelName, testModels, testData, votesRequired)
    }

    new KFoldsTrainingModel(
      labelName,
      mlModels.toList,
      votesRequired)
  }
```

TOYOTA

```
private def predictAll(
  labelName: String,
  mlModels: ParSeq[ClassificationModel],
  testData: RDD[LabeledPoint],
  votesRequired: Double,
```

Ver 1

56%

Accuracy

TOYOTA

Ver 3

36%

Accuracy

TOYOTA

Ver 8

35%

Accuracy

TOYOTA

False Positives

I just had my friend at the toyota dealer rotate my tires and he said … that the **brake** pads are getting thin really fast.
So what should I do when they get too thin in the future and start to **squeak**?

TOYOTA

False Positives

i cut the iac hose as shown in figure 20 in the manual but when i start the car, it started gasping for air... **choking**... **sounds** like it's about to die out.

i bought the power **brake** check valve (80190 part for kragen)... but either i'm not installing it right or it's the wrong size... i have no idea.

TOYOTA

Explicit Semantic Analysis

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information

Article Talk Read Edit View history Search

Drum brake

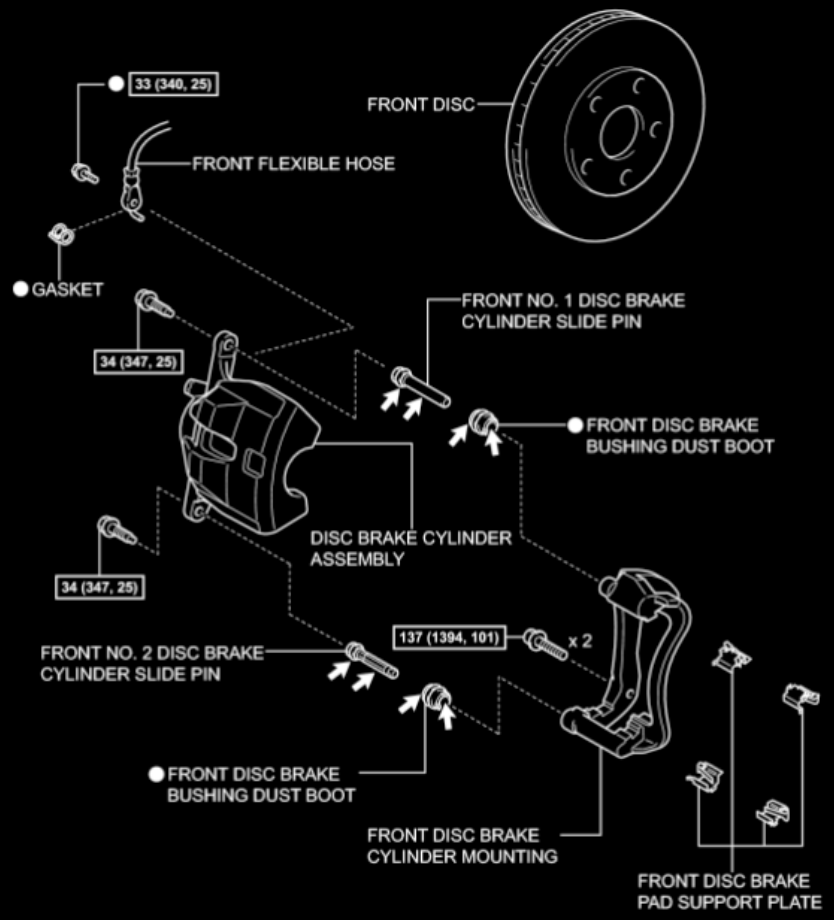
From Wikipedia, the free encyclopedia

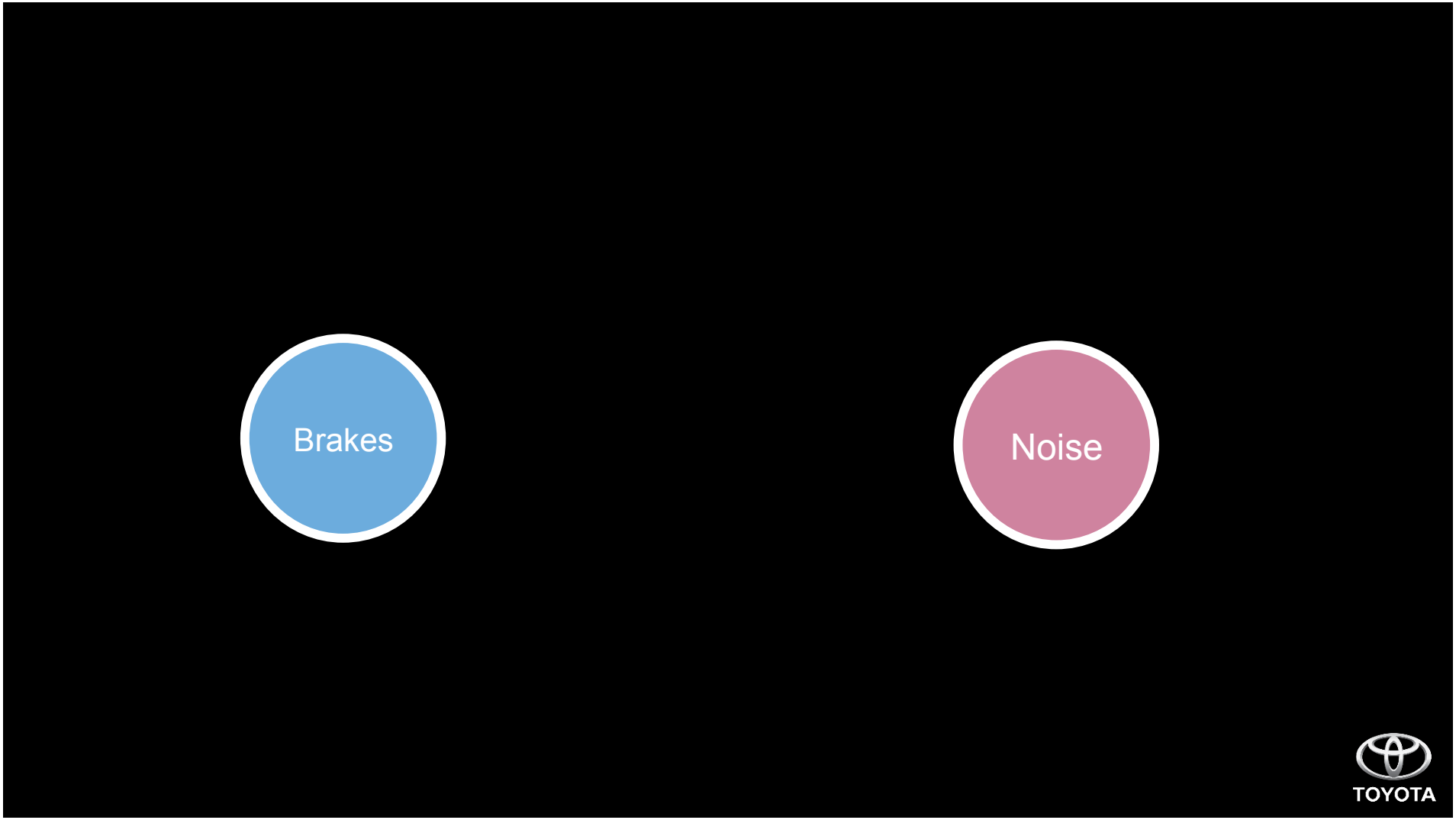
This article **relies largely or entirely upon a single source**. Relevant discussion may be found on the talk page. Please help improve this article by introducing citations to additional sources. *(July 2008)*

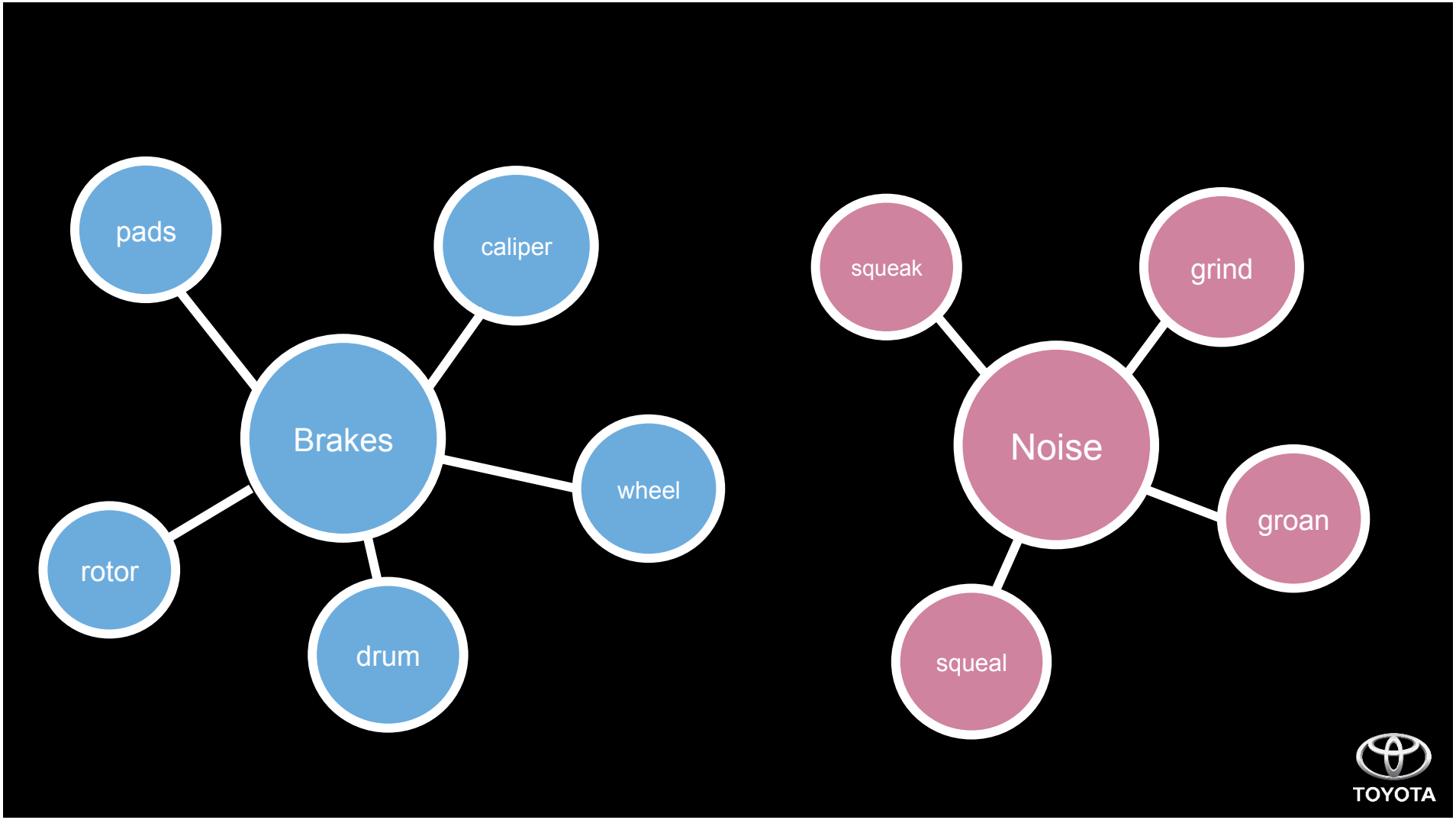
A **drum brake** is a brake that uses friction caused by a set of shoes or pads that press against a rotating drum-shaped part called a brake drum.

The term *drum brake* usually means a brake in which shoes press on the inner surface of the drum. When shoes press on the outside of the drum, it is usually called a *clasp brake*. Where the drum is pinched between two shoes, similar to a conventional disc brake, it is sometimes called a *pinch drum brake*, though such brakes are relatively rare. A related type called a band brake uses a flexible belt or

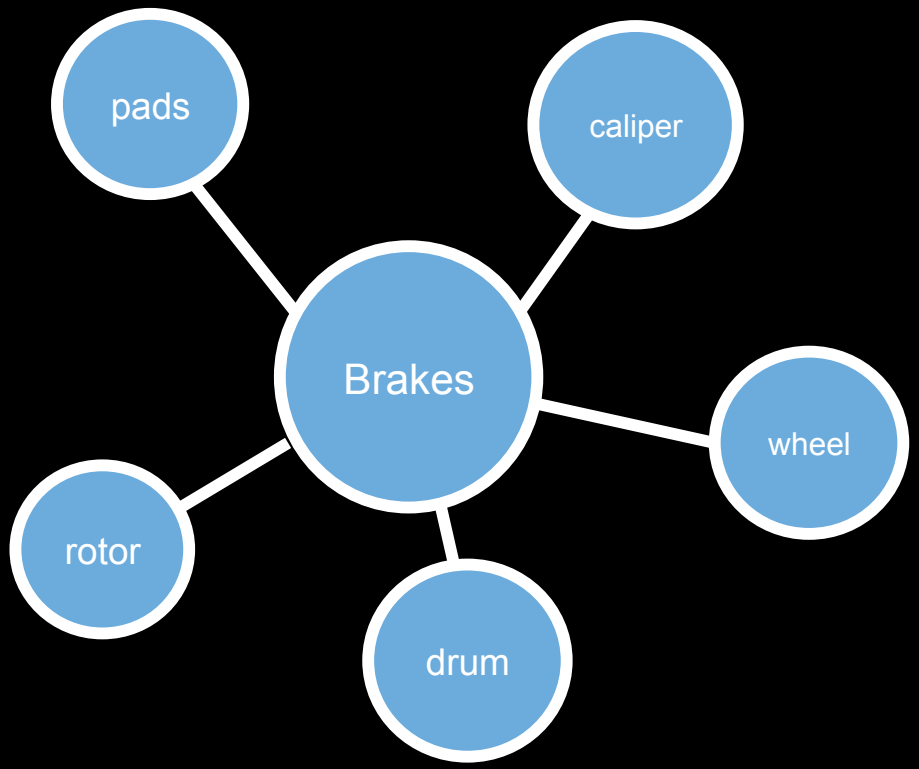
A drum brake with the drum removed, as used on the rear wheel of a car or truck. In this installation, a cable-operated parking brake operates the service shoes.



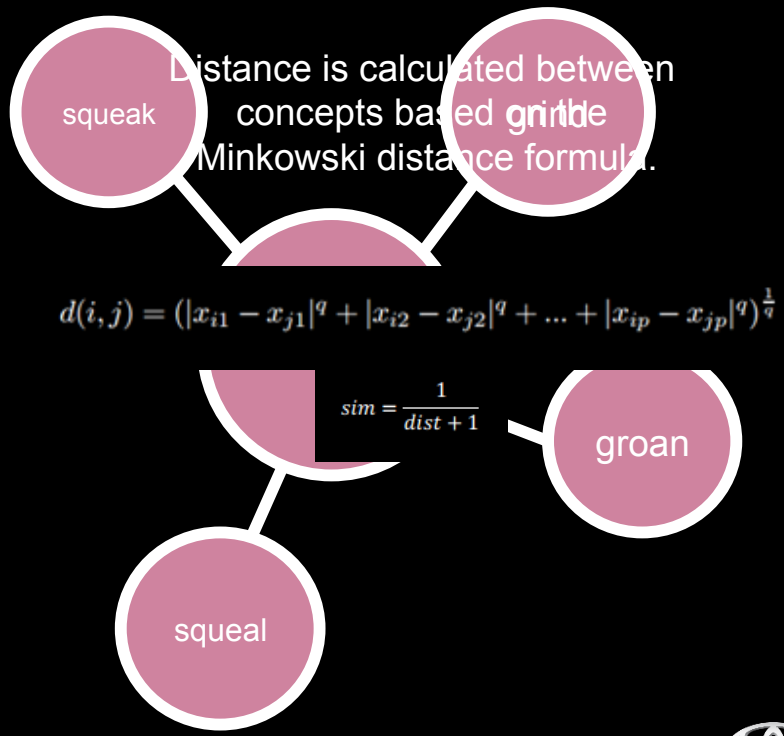

Brakes
Noise
TOYOTA


pads
caliper
Brakes
wheel
rotor
drum
squeak
grind
Noise
groan
squeal
TOYOTA

Distance Similarity Between Concepts



Distance is calculated between concepts based on the Minkowski distance formula.



$$d(i,j) = (|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \ldots + |x_{ip} - x_{jp}|^q)^{\frac{1}{q}}$$

$$sim = \frac{1}{dist + 1}$$

TOYOTA

Ver 9

82%

Accuracy

TOYOTA

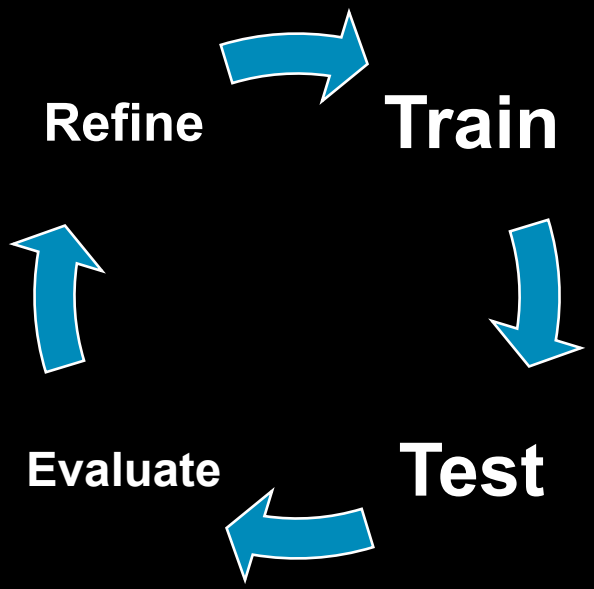
TOYOTA

Kaizen

改善 = Continuous Improvement

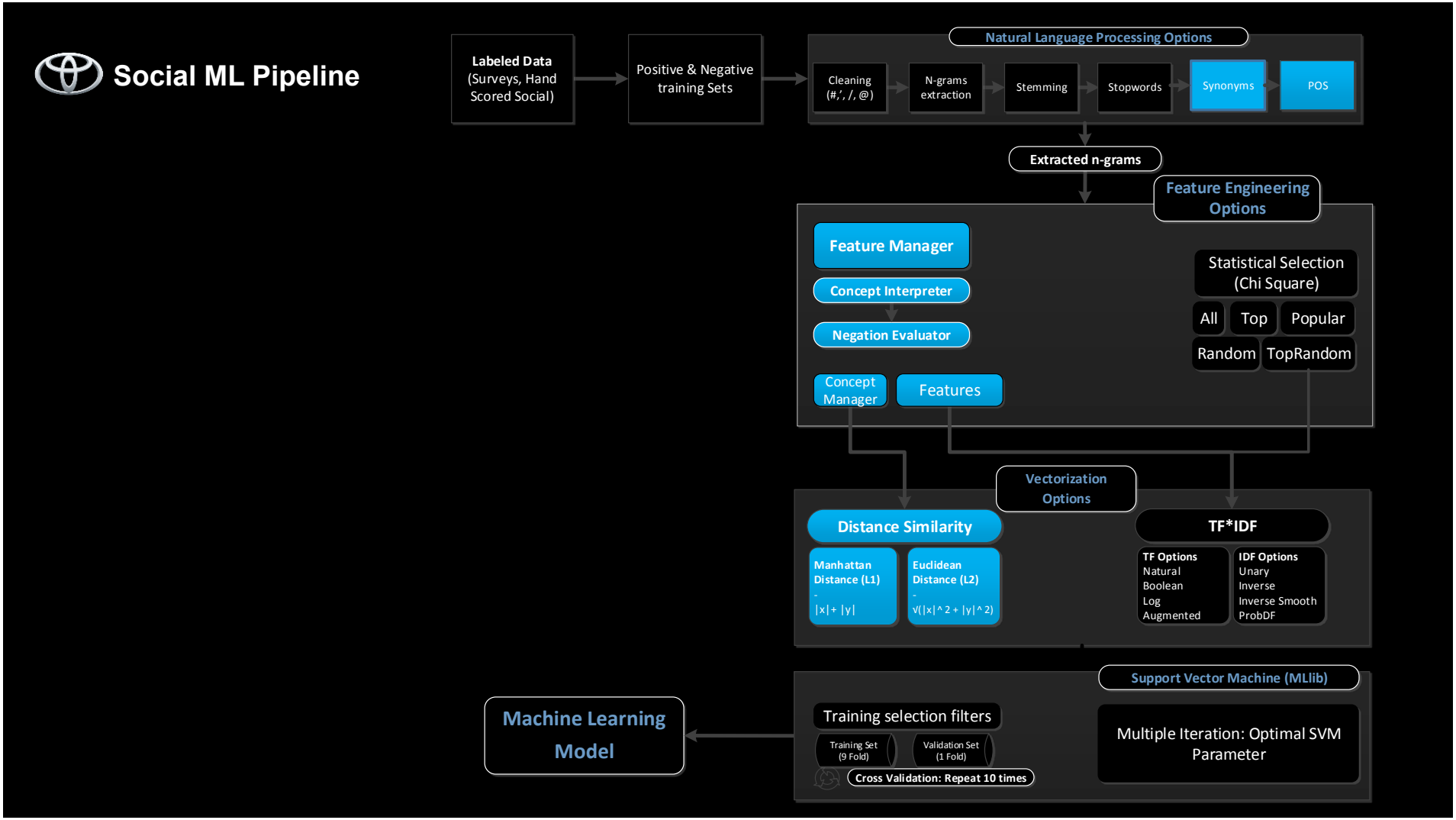
TOYOTA

Kaizen

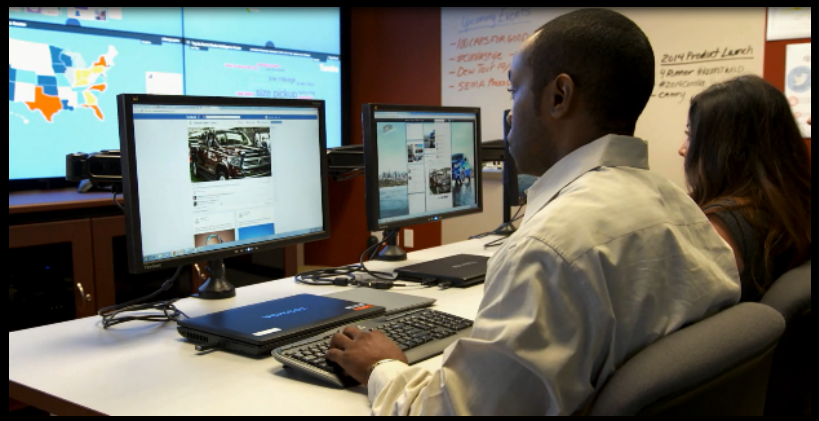


TOYOTA

Social ML Pipeline



TODAY



TOYOTA

FUTURE

Connected Vehicle Data

Consumer Data

Manufacturing Data

TEAM TOYOTA Spark Tips

- Education and Inclusion
- Pace Yourself
  - Design and Plan a Transitional Architecture to Incrementally Introduce Spark elements into your Applications
- Use Joda Time for Date Comparisons

TOYOTA

TEAM TOYOTA Lessons Learned

- Be mindful of AKKA versions when trying to Build a new Spark release to a packaged Hadoop Distribution
- Use SparkSQL versus DSLs for Joins
- Remember to configure Memory Fraction based on the size of your data.

TOYOTA

TOYOTA | Let's Go Places

Visit us at the **TOYOTA** Booth here at the Spark Summit Today.

Brian Kursar – Sr Data Scientist
Toyota Motor Sales IT Research and Development
@briankursar