

Analytics Concepts in Modern Age



A brief on Statistics Introduction to Analytics Machine Learning Artificial Intelligence

Data & Analytics

"The price of lightis less than the cost of darkness"

Arthur C. Nielsen

"A point of view can be a dangerous luxury when substituted for insight and understanding."

Marshall McLuhan,

"What gets measured gets managed" Peter Drucker

"If we have data, let's look at data. If all we have are opinions, let's go with mine" – Jim Barksdale

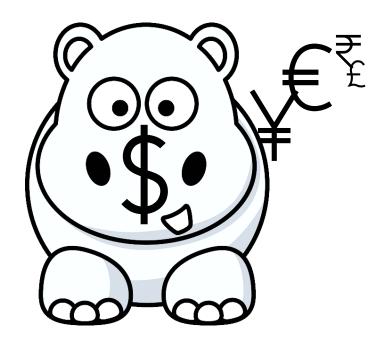
"In God we trust, all others must bring data". Edwards Deming

"We are drowning in information and starving for knowledge." – Rutherford D. Rogers

"Data beats emotions." Sean Rad,



No Analytics – Be ready for Hippo



Hippo – Highest Paid Person Opinion



STATISTCS – A Brief History

Basic forms of statistics have been used since the beginning of civilization.



Ancient civilizations used numbers for taking monitoring control over armies, food supplies and gauging strength of enemy armies

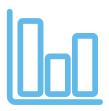


STATISTICS- A Definition

Statistics is a branch of mathematics dealing with data collection, organization, analysis, interpretation and presentation.

In applying statistics to, for example, a scientific, industrial, or social problem, it is conventional to begin with a statistical population or a statistical model process to be studied.











Father of Statistics

Sir Ronald Aylmer Fisher] (17 February 1890 – 29 July 1962) was a British statistician and geneticist. For his work in statistics, he has been described as "a genius who almost single-handedly created the foundations for modern statistical science" and "the single most important figure in 20th century statistics"

His contributions to statistics include the maximum likelihood, fiducial inference, the derivation of various sampling distributions, founding principles of the design of experiments, and much more.





One More Father

Prasanta Chandra Mahalanobis (29 June 1893 – 28 June 1972). He founded the Indian Statistical Institute, and contributed to the design of large-scale sample surveys.

For his contributions, Mahalanobis has been considered the **father of modern statistics in India**.

He is best remembered for the Mahalanobis distance, a statistical measure, and for being one of the members of the first Planning Commission of free India. He made pioneering studies in anthropometry in India





Scope of Statistical Theory

Scope

- Data collection
- Summarizing data
- Interpreting data
- Modelling
- Applied statistical inference



Operational Definitions Of Terms for Data

VARIABLE

A characteristic of an item or individual.

DATA

The set of individual values associated with a variable.

STATISTICS

The methods that help transform data into useful information for decision makers.



Types of Data

There are four types of data or levels of measurement:

CATEGORICAL OR NUMERIC

1. Nominal PAN, ID, ROLL NO.	2. Ordinal S&P Ratings, Grades
3. Interval Likert Scale, Temperature	4. Ratio Height, Weight

NUMERIC



Overview of Statistical Models

Type of Predictors Type of Response	Categorical	Continuous	Continuous and Categorical
Continuous	Analysis of Variance (ANOVA)	Ordinary Least Squares (OLS) Regression	Analysis of Covariance (ANCOVA)
Categorical	Contingency Table Analysis or Logistic Regression	Logistic Regression	Logistic Regression



Statistical Models

A statistical model is a mathematical model that embodies a set of statistical assumptions concerning the generation of sample data (and similar data from a larger population).

A statistical model represents, often in considerably idealized form, the datagenerating process

Unlike Movie Models



Statistical/Analytical/Decison Models

<u>Model in General</u>:

- ▶ An abstraction or representation of a real system, idea, or object
- ▶ Captures the most important features
- Can be a written or verbal description, a visual display, a mathematical formula, or a spreadsheet representation
- Largely these are equations used in Machine Learning, Al and Many More



Statistical Models



Figure 1.3



Statistical a.k.a Decision Models

- A <u>decision model</u> is a model used to understand, analyze, or facilitate decision making.
- ▶ Types of model <u>input</u>
 - data
 - uncontrollable variables
 - decision variables (controllable)



Decision Models

Descriptive Decision Models

- ▶ Simply tell "what is" and describe relationships
- ▶ Do not tell managers what to do



Dependent Variable (Y) & Independent Variable (Xs) Explained Variables (Y) & Explanatory Variables (Xs) Predicted Variables (Y) & Predictor Variables (Xs) Output Variables (Y) & Input Variables (Xs)

$$Y = b_0 + b_1x_1 + b_2x_2 + + b_nx_n$$

Where Y = Dependent Variable(DV)

 $X_1, x_2, x_n - Indepdent Variable(IV)$

b₀ - intercept

 b_1 , b_2 – coefficients

n – No. of observations



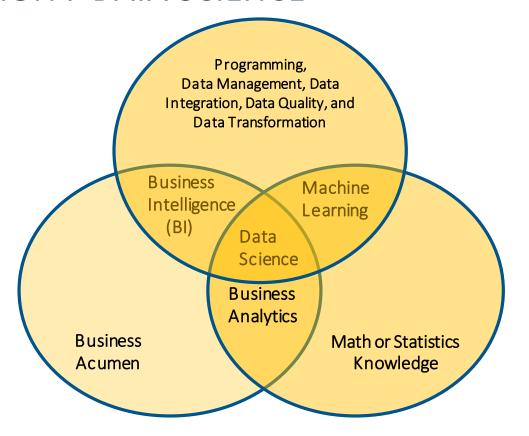
A brief on Statistics Introduction to Analytics Machine Learning Artificial Intelligence

TIME TO INTRODUCE ANALYTICS



SAS DEFINITION: DATA SCIENCE

SAS defines "data science" as "a multi-disciplinary field that combines skills in software engineering and statistics with domain experience to support the end-to-end analysis of large and diverse data sets, ultimately uncovering value for an organization and then communicating that value to stakeholders as actionable results."





TIME TO INTRODUCE: Analytics! What is Data Analytics?

Analytics is the use of:

```
data,
```

information technology,

statistical analysis,

quantitative methods, and

mathematical or computer-based models

to help managers gain improved insight about their business operations and make better, fact-based decisions.

Business Analytics (BI) is a subset of **Data Analytics**



What is Business Analytics?

Business Analytics Applications

- Management of customer relationships
- Financial and marketing activities
- Supply chain management
- ▶ Human resource planning
- Pricing decisions
- ▶ Sport team game strategies



What is Business Analytics?

<u>Importance of Business Analytics</u>

- ▶ There is a strong relationship of BA with:
 - profitability of businesses
 - revenue of businesses
 - shareholder return
- ▶ BA enhances understanding of data
- ▶ BA is vital for businesses to remain competitive
- ▶ BA enables creation of informative reports



Scope of Business Analytics

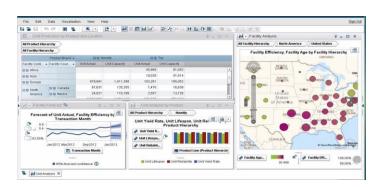
- Descriptive analytics
 - uses data to understand past and present
- Predictive analytics
 - analyzes past performance
- Prescriptive analytics
 - uses optimization techniques



Descriptive Analytics

- Descriptive analytics, such as reporting/OLAP, dashboards, and data visualization, have been widely used for some time.
- They are the core of traditional BI.

Variable	N	Mean	Std Dev	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
MSRP	428	32774.86	19431.72	10280.00	20329.50	27635.00	39215.00	192465.00
Invoice	428	30014.70	17642.12	9875.00	18851.00	25294.50	35732.50	173560.00
EngineSize	428	3.1967290	1.1085947	1.3000000	2.3500000	3.0000000	3.9000000	8.3000000
Cylinders	426	5.8075117	1.5584426	3.0000000	4.0000000	6.0000000	6.0000000	12.0000000
Horsepower	428	215.8855140	71.8360316	73.0000000	165.0000000	210.0000000	255.0000000	500.0000000
MPG_City	428	20.0607477	5.2382176	10.0000000	17.0000000	19.0000000	21.5000000	60.0000000
MPG_Highway	428	26.8434579	5.7412007	12.0000000	24.0000000	26.0000000	29.0000000	66.0000000
Weight	428	3577.95	758.9832146	1850.00	3103.00	3474.50	3978.50	7190.00
Wheelbase	428	108.1542056	8.3118130	89.0000000	103.0000000	107.0000000	112.0000000	144.0000000
Length	428	186.3621495	14.3579913	143.0000000	178.0000000	187.0000000	194.0000000	238.0000000





What has occurred?

Descriptive analytics, such as data visualization, is important in helping users interpret the output from predictive and predictive analytics.

Decision Models

A Linear Demand Prediction Model

As price increases, demand falls.





Decision Models

A Nonlinear Demand Prediction Model

Assumes price elasticity (constant ratio of % change in demand to % change in price)

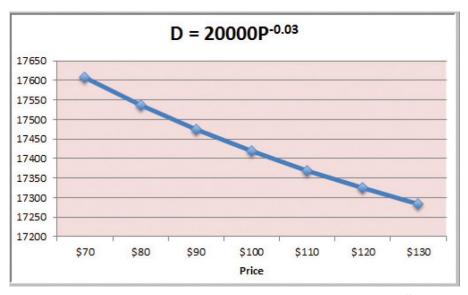


Figure 1.9



Dependent Variable (Y) & Independent Variable (Xs) Explained Variables (Y) & Explanatory Variables (Xs) Predicted Variables (Y) & Predictor Variables (Xs) Output Variables (Y) & Input Variables (Xs)

$$Y = b_0 + b_1x_1 + b_2x_2 + + b_nx_n$$

Where Y = Dependent Variable(DV)

 $X_1, x_2, x_n - Indepdent Variable(IV)$

b₀ - intercept

 b_1 , b_2 – coefficients

n – No. of observations



Example of Ys and Xs

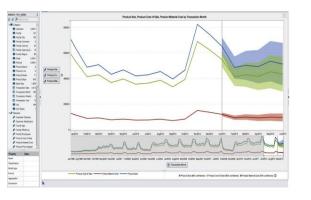
Y can be Sales, which one might want to predict

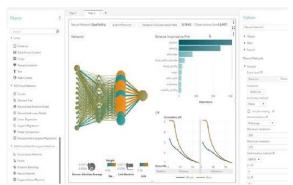
CAN YOU THINK MORE!! WHAT MIGHT AFFECT SALES ??

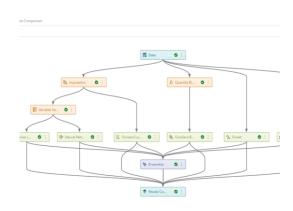


Predictive Analytics

- Algorithms for predictive analytics, such as **regression analysis**, **machine learning**, **and neural networks**, have also been around for some time.
- Prescriptive analytics are often referred to as advanced analytics.







What will occur?

- Marketing is the target for many predictive analytics applications.
- Descriptive analytics, such as data visualization, is important in helping users interpret the output from predictive and prescriptive analytics

Decision Models

<u>Prescriptive Decision Models</u> help decision makers identify the best solution.

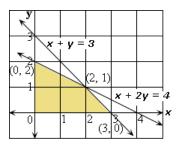
- Optimization finding values of decision variables that minimize (or maximize) something such as cost (or profit).
- ▶ <u>Objective function</u> the equation that minimizes (or maximizes) the quantity of interest.
- Constraints limitations or restrictions.
- Optimal solution values of the decision variables at the minimum (or maximum) point.



Prescriptive Analytics

- Prescriptive analytics is most often related to optimization
 - Often for the allocation of scarce resources

```
The OPTMODEL Procedure
                Problem Summary
    Objective Sense
                               Minimization
   Objective Function
   Objective Type
                                  Nonlinear
   Number of Variables
   Bounded Above
   Bounded Below
   Bounded Below and Above
   Fixed
   Number of Constraints
   Linear LE (<=)
   Linear EQ (=)
   Linear GE (>=)
   Linear Range
            The OPTMODEL Procedure
               Solution Summary
                             NLPC/Trust Region
Objective Function
Solution Status
                                       Optimal
Objective Value
                                          -3300
Iterations
Absolute Optimality Error
Relative Optimality Error
Absolute Infeasibility
                                  3.552714E-15
Relative Infeasibility
                                  4.866731E-17
                          11
```



What should occur? What to do?

- For example, the use of linear programming for yield management is common for organizations that have "perishable" goods (e.g., rental cars, hotel rooms, airline seats).
- Airlines have been using revenue management for pricing for a long time.



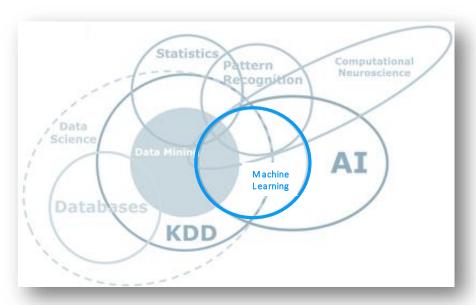
A brief on Statistics Introduction to Analytics Machine Learning Artificial Intelligence

Break for 15 mins We shall resume by 10.45 am



What is Machine learning?





SAS Data Mining Primer course 1998

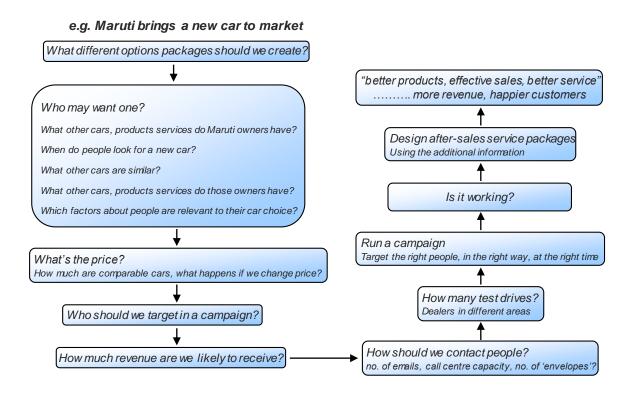
Machine learning is a branch of artificial intelligence that automates the building of systems that learn iteratively from data, identify patterns, and predict future results — with minimal human intervention.

It shares many approaches with other related field, but it focuses on predictive accuracy rather than interpretability of the model

FUN FACT: More than 30 years ago, SAS CEO, Jim Goodnight wrote a procedure for "k-nearest neighbor discriminant analysis," which is a machine learning method! And growing since....



Machine learning helps in business process





Machine learning helps in business process Welcome to the Era of Machine e.g. Maruti bring a new car to market Learning! experimental design LLL V visualisation a.dll associations experimental design associations: sequence segmentation associations monitor variable identification <u>©</u> deployment simulation forecasting event based modelling value prediction constraint based optimisation



Machine Learning

SUPERVISED LEARNING

Regression

LASSO regression Logistic regression Ridge regression

Decision tree

- Gradient boosting Random forests
- Neural

networks

- SVM
- Naïve Bayes Neighbors Gaussian processes

UNSUPERVISED **LEARNING**

A priori rules

Clustering

Lk-means clustering

Mean shift clustering

Spectral clustering

Kernel density

- estimation Nonnegative
- matrix factorization PCA

Kernel PCA Sparse PCA

Singular value decomposition SOM

SEMI-**SUPERVISED** LEARNING

Prediction and classification*

- Clustering*
- EM **TSVM**
- Manifold regularization Autoencoders
- Multilayer perceptron

Restricted Boltzmann machines



TRANSDUCTION

REINFORCEMEN **LEARNING**

DEVELOPMENTA LEARNING

*In semi-supervised learning, supervised prediction and classification algorithms are often combined with clustering.



Machine learning: Why is it so important now?

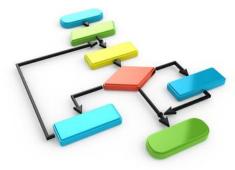












Data



Computing Power

Data

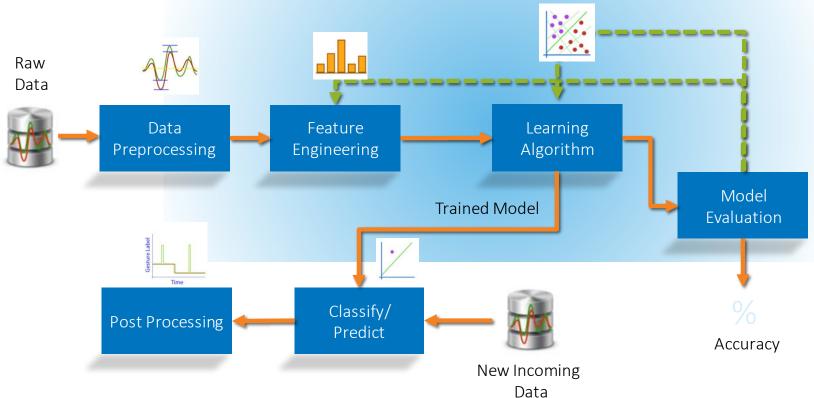






Machine learning process flow





Applications of machine learning

Credit Scoring
Next Best Offers



Fraud



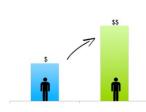
Customer Segmentation



Online Recommendations



Customer Lifetime Value



Targeted Acquisition / Retention / Attrition



Real-time Ad placements



Natural Language Processing



Network Intrusion
Detection



Predictive Asset Maintenance





A brief on Statistics Introduction to Analytics Machine Learning Artificial Intelligence



You Are ...

Leaving for a business trip tomorrow?

 Your intelligent device will automatically offer weather reports and travel alerts for your destination city.



You Are ...

Leaving for a business trip tomorrow?

 Your intelligent device will automatically offer weather reports and travel alerts for your destination city.

Planning a large birthday celebration?

 Your smart bot will help with invitations, make reservations and remind you to pick up the cake.





You Are ...

Leaving for a business trip tomorrow?

 Your intelligent device will automatically offer weather reports and travel alerts for your destination city.

Planning a large birthday celebration?

 Your smart bot will help with invitations, make reservations and remind you to pick up the cake.



Planning a direct marketing campaign?

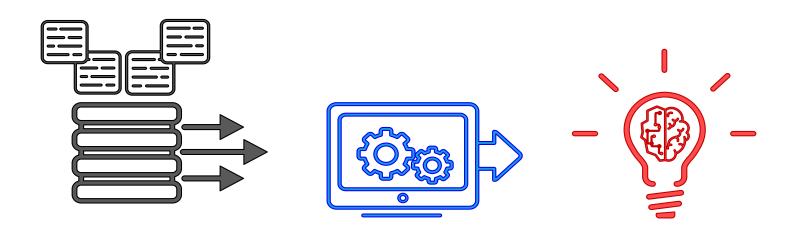
 Your Al assistant can instinctively segment your customers into groups for targeted messaging and increased response rates.





Artificial Intelligence (AI)

Artificial intelligence (AI) is the science of training computers to perform tasks that typically require human intelligence to complete.





Artificial Intelligence

is the science of training systems to emulate human tasks through

Learning and **Automation**







Evolution of Artificial Intelligence



1950s-1970s Neural Networks



1980s-2010s Machine Learning

SAS



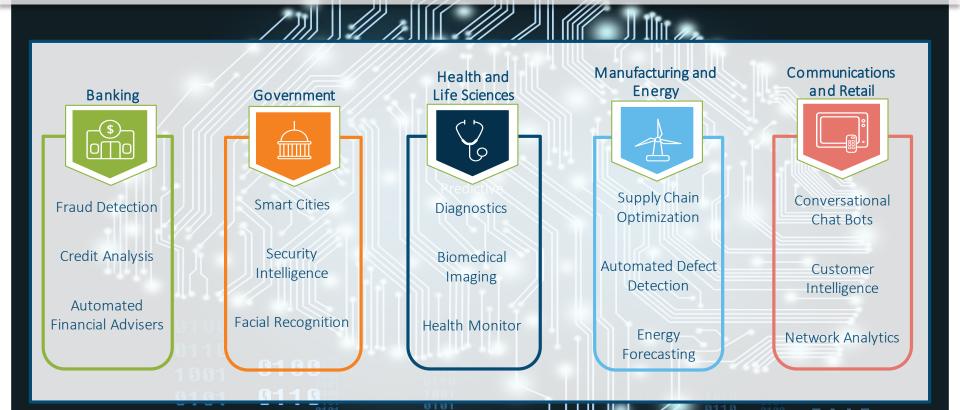
Present Day
Deep Learning and
Cognitive Systems

1976

)101 **0110** |100 **1001**

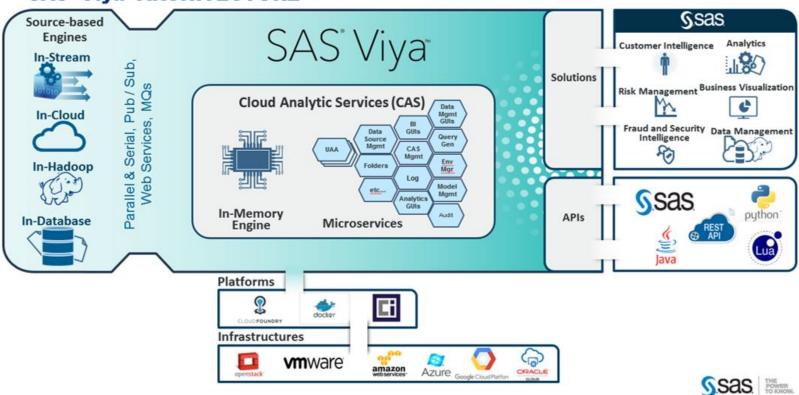


Applications of Artificial Intelligence

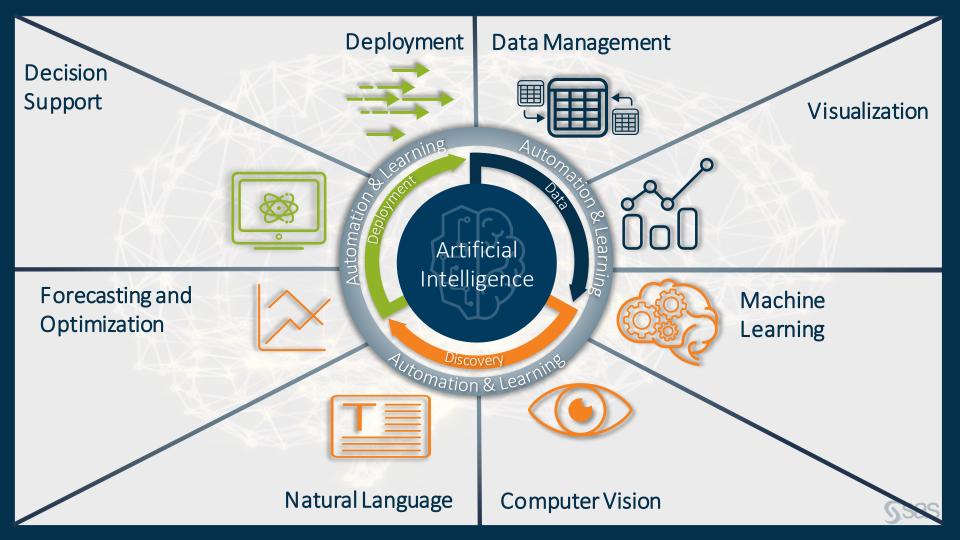


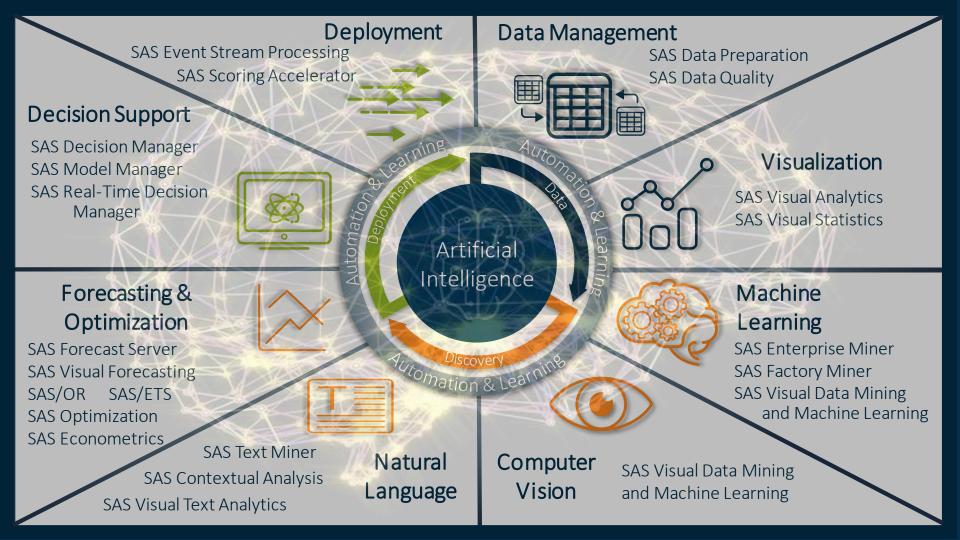
0101

SAS° Viya" ARCHITECTURE









What is Deep Learning?

Deep Learning is a type of machine learning used in recognizing speech, identifying objects in images and more.



Example of Classification: Images

Problem: Build a Model to classify "Images" into groups









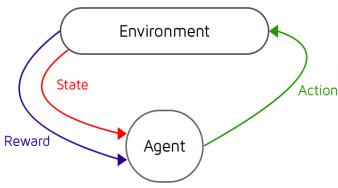






Reinforcement Learning

 Discovers through trial and error which actions yield the greatest rewards.





 Agent choose actions that maximize the expected reward over a given amount of time. The agent will reach the goal much faster by following a good policy.



Often used for robotics, gaming and navigation.





Computer Vision makes it possible to spot defects not easily visible to the human eve











Elon Musk warns that an 'immortal' digital dictator could forever trap humanity in its grasp unless we start regulating technology ASAP.

(https://www.livescience.com/62239-elon-musk-immortal-artificial-intelligence-dictator.html)

