Business Analytics Research and Teaching Perspectives

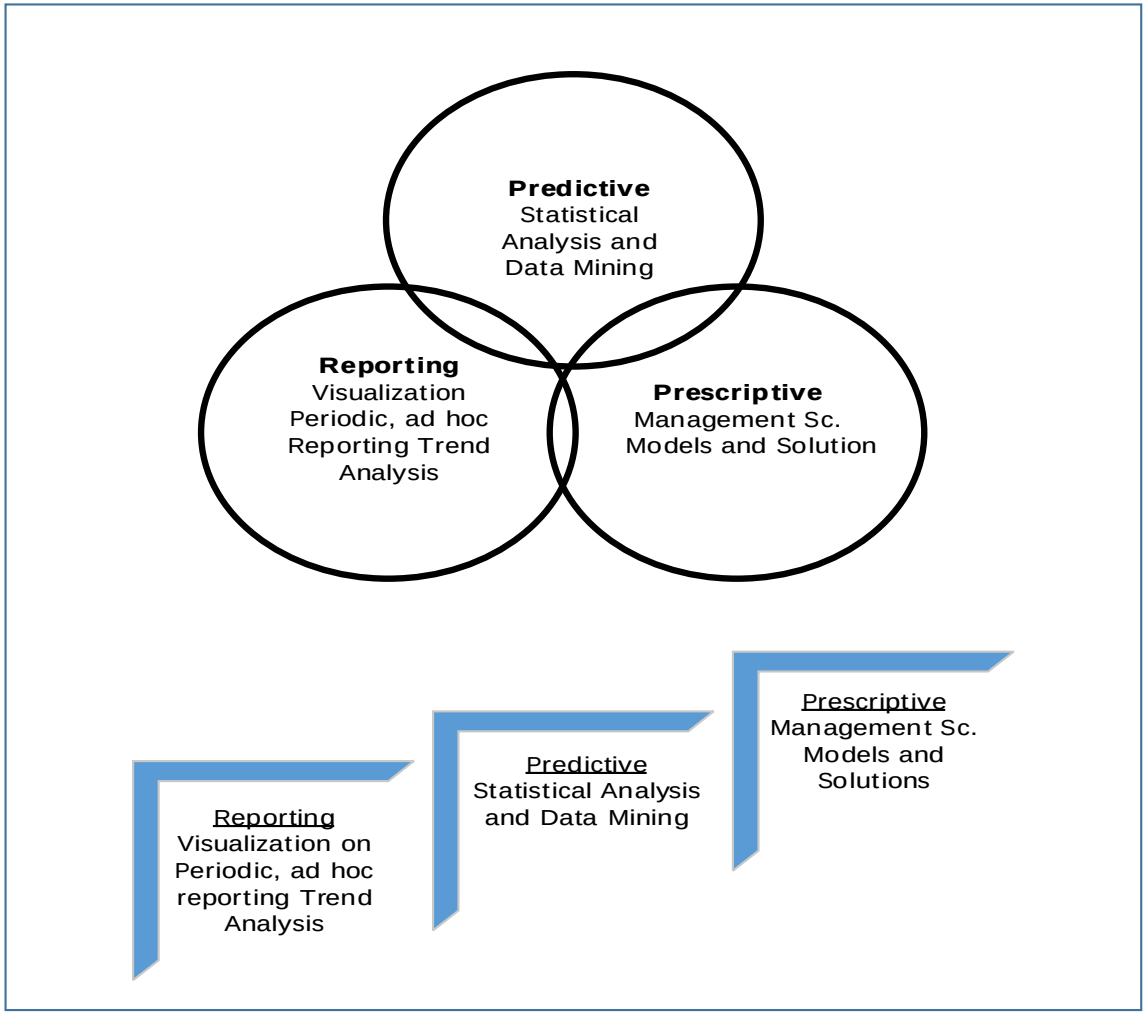
Ramesh Sharda, Daniel A. Asamoah and Natraj Ponna

Institute for Research in Information Systems,
Spears School of Business,
Oklahoma State University
Stillwater, OK 74078

Outline

- Three Types of Analytics
- Research Perspectives
- Athletic Injury Analysis
- Big Data Research
- Teaching Perspectives
- Analytics Ecosystem

Types of Analytics



Research Perspectives

- Significant growth in all three types of analytics
- Prescriptive analytics - Application of ORMS techniques in building revenue optimization models
- Predictive analytics – Development of algorithms and applying them in new and innovative settings
- Big Data analytics – volume, variety, velocity of data
- Current opportunities – sensors/large scale observations/analyses
- Examples
  - Entertainment industry
  - Healthcare
  - Sports

Predicting Injury Recovery Time in College Football

Daniel Asamoah and Ramesh Sharda

Research Background

- NCAA was formed to help curb occurrence of college football injuries and deaths
  - 66,313 college students participated in NCAA's Divisions I, II and III football leagues in 2009-2010 season
  - Football is still one of the most injury-prone sports

- Aim was to investigate injury occurrences and PREDICT how long an injury would take to heal

Introduction: Why Injury Recovery Time?

- Despite efforts by NCAA, there is still frequent occurrence of injuries
- Uncertainty of how long an injury might take to heal
- This affects a coach's decision on player management and tactics. Knowledge of injury recovery time could be used in a decision support system by coaches and medical staff to manage injuries and plan future game strategies with cognizance of how long players would take to recover from injuries

Sports Injury Studies

- Most occurring injuries (Powell et al. 1999; Vastag, 2002)
- Ways of managing football injuries for overall team performance (Fukada et al., 2012; Dick et al., 2007)
- Injury Analysis (Feeley et al., 2008; Comin et al., 2012)
- Re-injury prevention (Heiderscheit et al., 2010)
- Post-injury analysis of certain conditions such as concussion (McCrea et al., 2003)
- Variables that affect injury recovery (Lau et al., 2011; Tuberville et al.; 2003)

Research Gap

- Studies that have investigated injury recovery have mostly used only traditional regression or descriptive statistical analysis methods.

- Focus has been on factors that affect injury recovery time rather than make quantifiable time prediction for recovery.

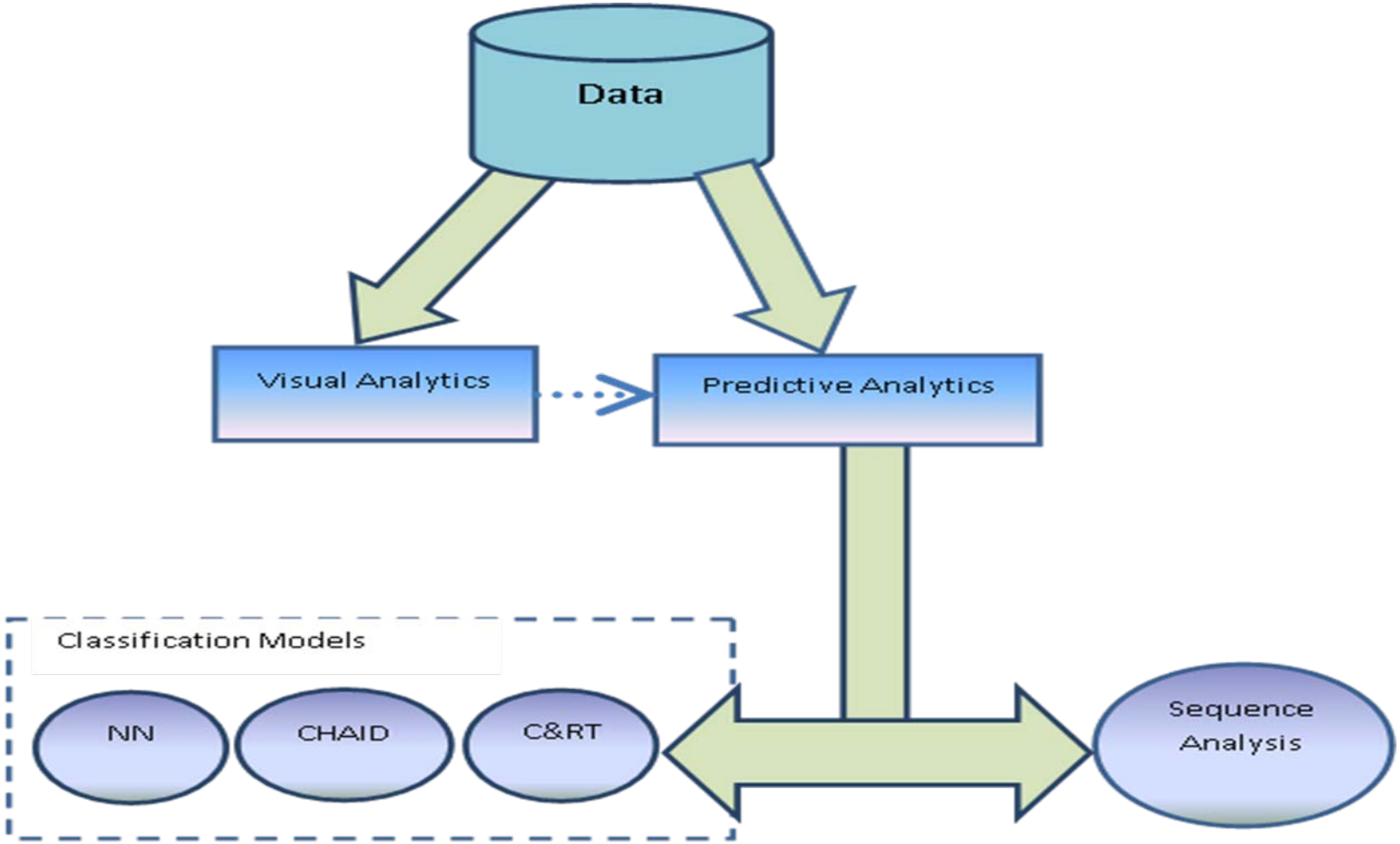
Research Question

- How accurately can college football injury recovery time be predicted using data mining methods?
  - what are the most important factors that affect occurrence of injuries?

Methodology: Data Mining

- Data mining is the process of analyzing data from different perspectives and summarizing it into useful information
  - Visual Analytics
  - Predictive Analytics
  - Decision Tree: data classification and prediction method commonly used due to its intuitive ability to explain relationship among several variables.
  - Neural Network (NN): a mathematical and computational model for pattern recognition and data classification through a learning process.
  - Sequence Analysis: a method for understanding the relationships among a group of variables based on the order of occurrence.

Methodology: Data Mining



Methodology: Data Collection and Preparation

- Data was collected from a healthcare IT company
- Off-field data was deleted
- Variables were categorized as Intrinsic and Extrinsic variables
- Target variable (recovery time) was generated based on 'injury date' and 'resolved date' variables
- Total of 17 variables and 384 football records were left after data reduction.
- The data set was partitioned into two sub sets; a training set (70%) and a testing set(30%).

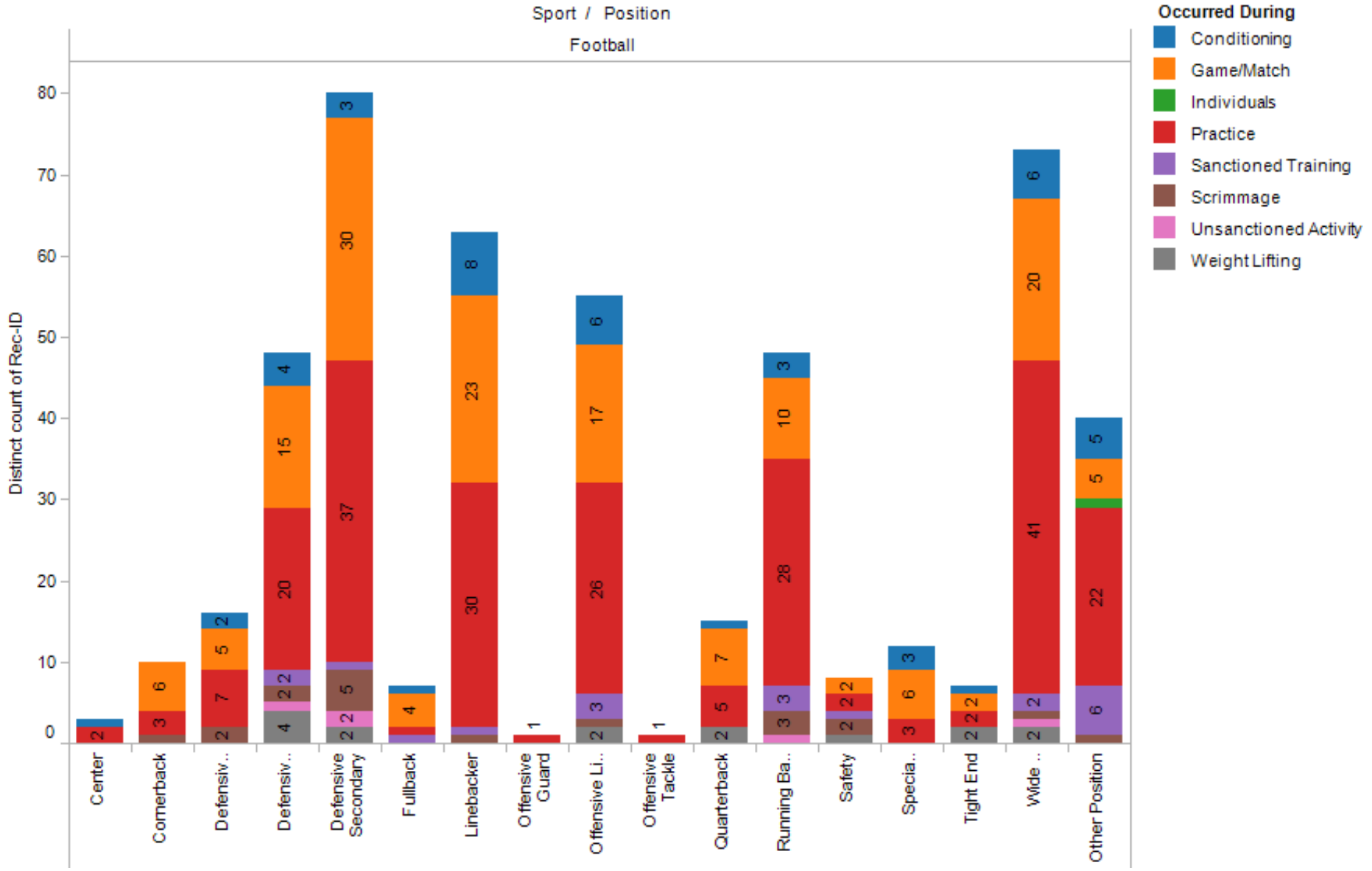
Variables Used

No.	Variable Name	Variable Description
1	RECID	Unique player identification
2	Position	Position athlete was playing when injury occurred
3	Injury Date	Date injury occurred
4	Closed Date	Date injury was resolved
5	Recovery Time	Period injury took to heal
6	Action Taken	Who handled injury when it occurred e.g. team physician
7	Body Part	Part of body where injury occurred e.g. knee
8	Laterality	Lateral position of injury on body e.g. Left knee
9	Injury	Type of injury that occurred e.g. strain
10	Severity	Degree of seriousness of injury
11	Current Status	Seriousness of injury on a progressive basis
12	Occurred During	Type of event when injury occurred e.g. practice session
13	Event Location	Location of field when injury occurred
14	Field Type	Type of field surface when injury occurred e.g. grass
15	Opponent	Opposing team when injury occurred
16	Activity	Motion athlete was in when injury occurred e.g. running
17	Onset	How injury happened e.g. gradual, acute contact

Results - Visualization

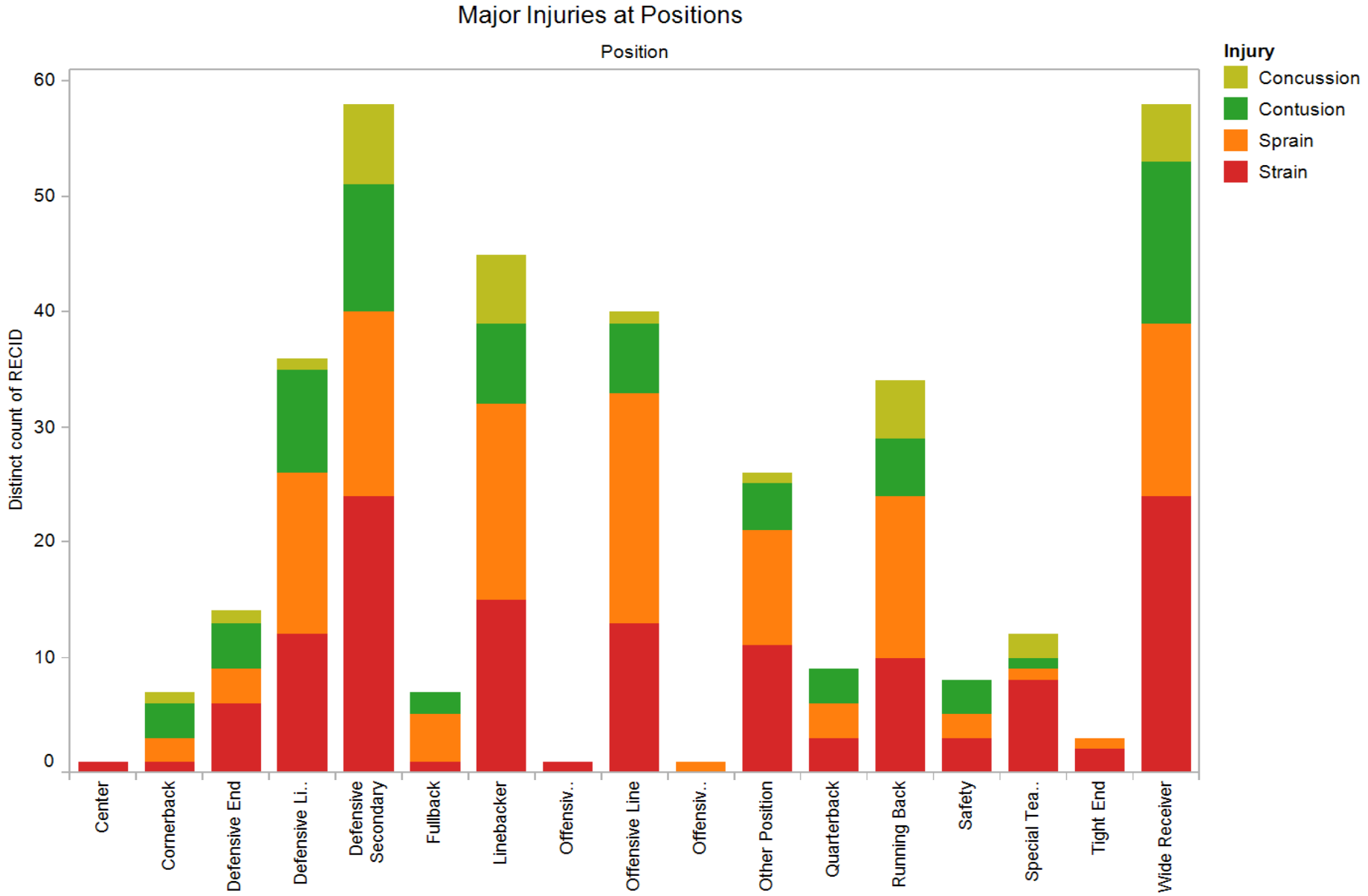
Injuries During Different Activities

- More injuries occurred during practice sessions than during actual games
  - Adequate safety mechanisms were probably not enforced during training sessions as compared to actual games.


Sport / Position
Football
Distinct count of Rec-ID
0
10
20
30
40
50
60
70
80
Center
Cornerback
Defensiv..
Defensiv..
Defensive Secondary
Fullback
Linebacker
Offensive Guard
Offensive Li..
Offensive Tackle
Quarterback
Running Ba..
Safety
Specia..
Tight End
Wide ..
Other Position
Occurred During
Conditioning
Game/Match
Individuals
Practice
Sanctioned Training
Scrimmage
Unsanctioned Activity
Weight Lifting

Major Injuries at Various Playing Positions

- Most frequently occurring injuries
  - Strain, Sprain, Contusion and Concussion in decreasing order of frequency.
- Players in defensive positions are more prone to injuries


Major Injuries at Positions
Position
Injury
Concussion
Contusion
Sprain
Strain
Distinct count of RECID
60
50
40
30
20
10
0
Center
Cornerback
Defensive End
Defensive Li..
Defensive Secondary
Fullback
Linebacker
Offensiv..
Offensive Line
Offensiv..
Other Position
Quarterback
Running Back
Safety
Special Tea..
Tight End
Wide Receiver

Results - Prediction Models

Prediction Models

- A NN model was built based on recovery time in months
- Consequently, for a more refined (and informative) model, NN, CHAID and C&RT models were built based on weeks
- Determined “Within 1 category away” - The accuracy of the prediction 1 step away from the correct prediction.

Model Evaluation

- Model 1(Neural Network Model)
  - Generally better when categories in output variable (recovery time) were few.
  - Predicted early recovery time much better.
- Model 2 (CHAID Model)
  - Fared better when categories in output variable (recovery time) was increased.
  - Predicted early and mid-term recovery time much better.
- Model 3 (C&RT)
  - Predicted late-term recovery time much better
  - Improved model accuracy in the 'within 1 category away' measure
- Model 4 (Sequence Analysis)
  - Portrayed interesting trends in the injury data

Neural Network Model Evaluation

Classification table for NN's performance (in months)

Observed \ Predicted	0-1Month	1-2Months	2-4Months	4-6Months	6-24Months
0-1Month	100.0%	0.0%	0.0%	0.0%	0.0%
1-2Months	0.0%	93.3%	3.3%	3.3%	0.0%
2-4Months	4.5%	0.0%	86.4%	9.1%	0.0%
4-6Months	0.0%	0.0%	22.2%	77.8%	0.0%
6-24Months	0.0%	20.0%	0.0%	10%	70.0%

Neural Network Model Evaluation

Accuracy measures for NN (in months)

Months to recovery / Predicted/ observed	0-1	1-2	2-4	4-6	6-24
Accuracy (%)	100	93.3	86.4	77.8	70
Within 1 category away (%)	100	96.7	95.5	100	80

Model Evaluation

Accuracy measures for NN, C&RT and CHAID (in weeks)

Predicted/observed \ Weeks to recovery		0-1	1-2	2-3	3-4	4-8	8-16	16-96
CHAID	Accuracy (%)	80	65.2	80.1	58.8	38.1	17.7	45.5
	Within 1 category away (%)	88.6	93.5	88.5	88.2	47.6	23.5	45.5
C&RT	Accuracy (%)	46.0	60.4	78.3	68.4	31.8	64.7	90.9
	Within 1 category away (%)	67.6	81.2	95.7	84.2	50.0	76.5	90.9
NN	Accuracy (%)	18.8	65.5	25.7	4.8	15.4	0.0	6.2
	Within 1 category away (%)	87.5	93.1	74.3	42.9	19.2	0.0	6.2

Sequence Analysis (Both Body Parts and Injuries)

- Injuries in lower extremities seem to result in higher extremity injuries
  - Those who have had Strain on Femur are likely to suffer injuries on shoulder and knee.
- Injuries seem to repeat
  - Those who have had ankle sprains are more likely to suffer another injury on ankle.
  - Precautionary measures could be used to minimize recurrence

Antecedent	Consequent	Support (%)	Confidence (%)
Femur and Strain	Shoulder	25.5	22.5
Femur and Strain	Knee	25.5	20.0
Ankle and Sprain	Ankle	22.3	25.7
Ankle and Sprain	Ankle	22.3	25.7
Ankle and Sprain	Hand	22.3	20.0
Femur and Strain	Femur	25.5	30.0

Recommendation

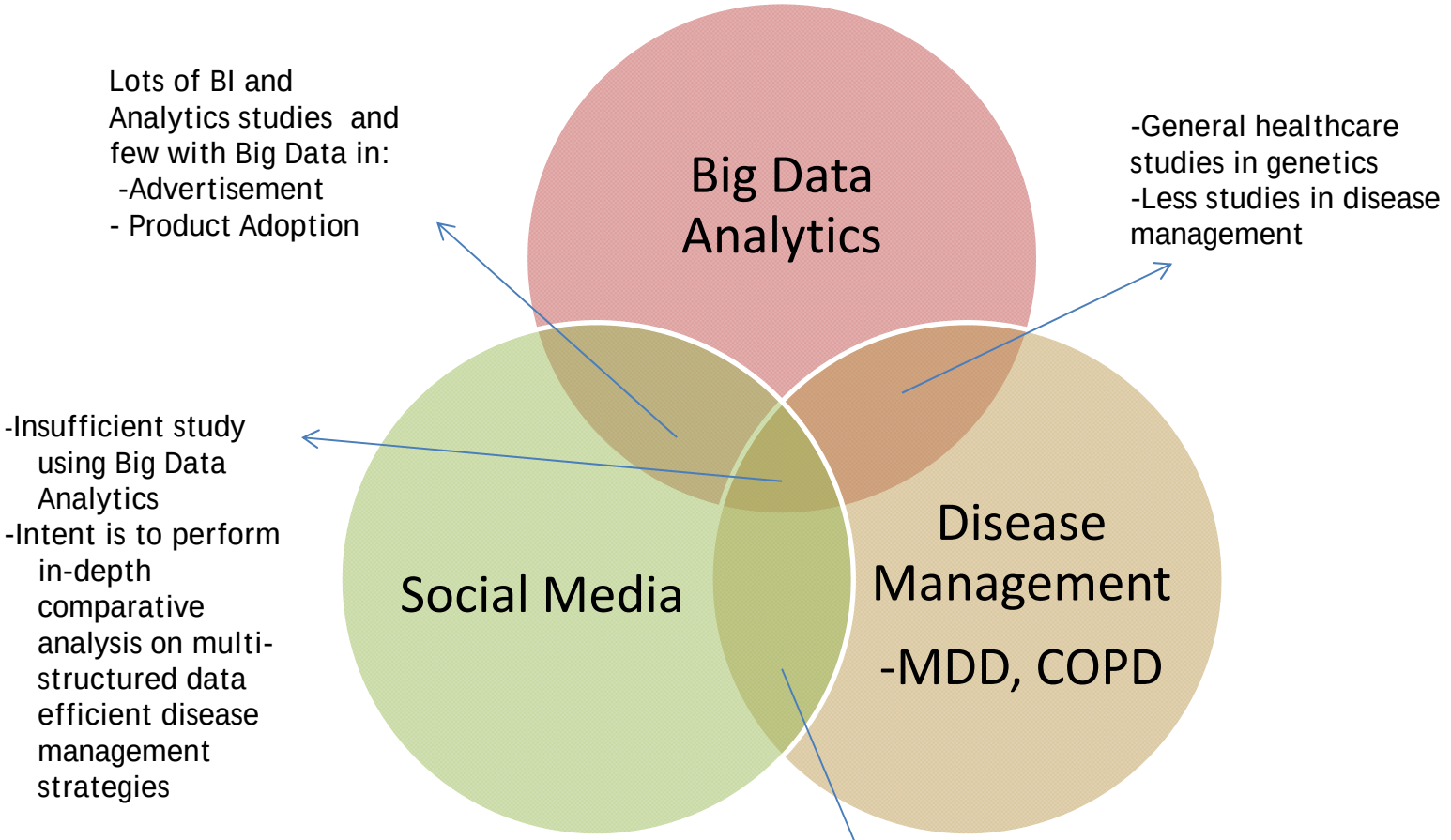
- Most injuries are treated by training room staff. To decrease injury time some of these might be taken to the specialist right from injury onset.
- Significant amount of injury occurred during practice. Adequate safety mechanisms have to be in place during training too.
- When a player is prone to a type of injury given past injury occurrence, precautionary measures should be put in place in order to minimize future injuries (based on sequence analysis).
- Players in defensive positions are more prone to injuries.

Big Data Research Opportunities

- Social Media + Healthcare
- Marketing – Twitter analysis
- Sensor data applications (RFID etc.)

Large Scale Mining of Social Media Healthcare Data: A Patient-Centered Perspective of Chronic Disease Management

Big Data Research Opportunities



DEPARTMENT OF MANAGEMENT SCIENCE & INFORMATION SYSTEMS
IRIS: INSTITUTE FOR RESEARCH IN INFORMATION SYSTEMS

Modeling Brand Post Popularity in Online Social Networks

Research Motivation

- *"60% of social customers recommend a brand to a friend after following the brand on Twitter or Facebook"*

- *"50% of them are more likely to buy from that brand as well (Constant Contact, 2011)"*

Research Questions

- *"How information moves on online social networks platforms?"*
- *"How users respond to various stimuli?"*
- *"How a content becomes popular on OSNs?"*

Research Objective

"Use Hawkes process as one of the most widely used point process in the literature to shed light on how the content popularity on OSNs can be described by **a function of time and the number of followers**"

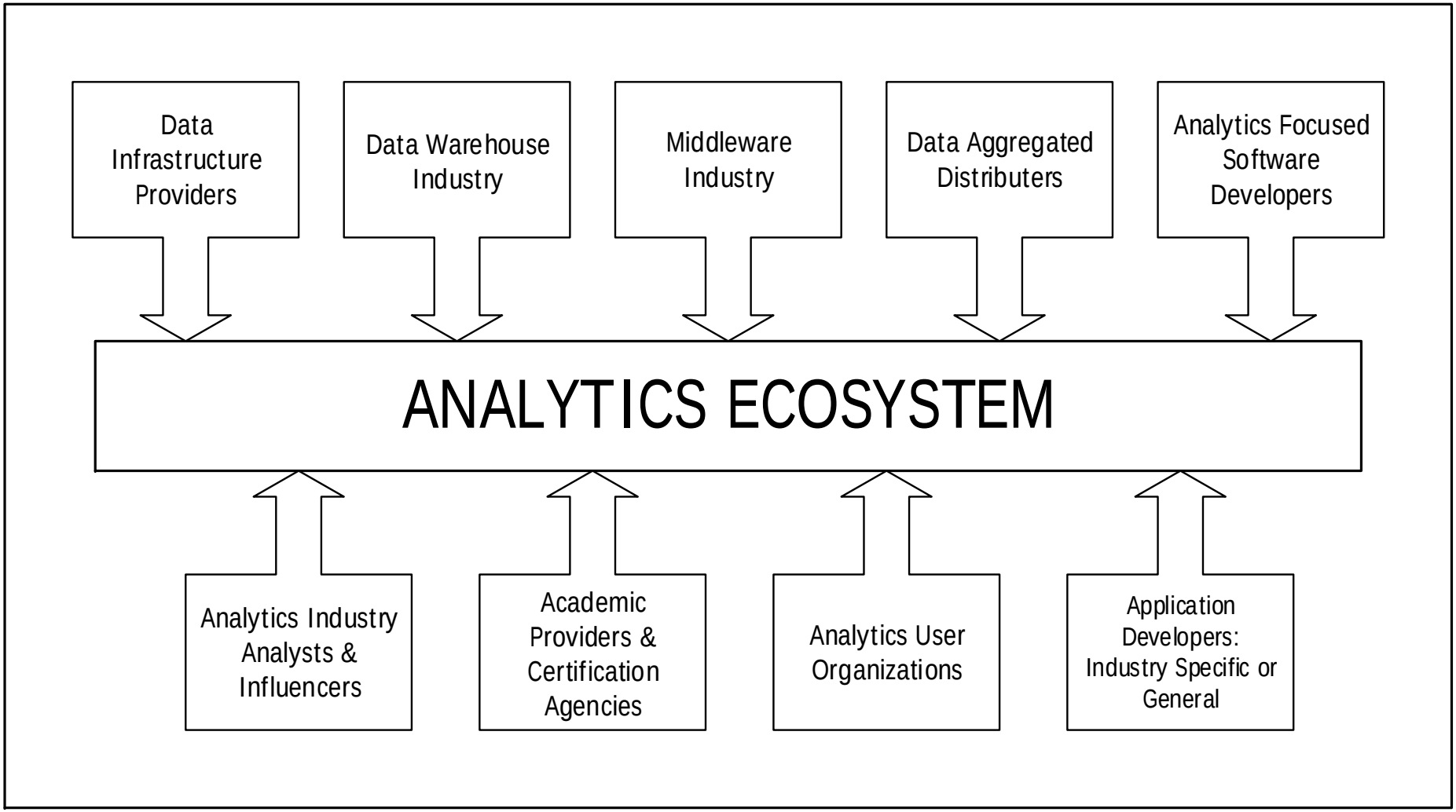
Managerial Implementations

- Influential users in amplifying the brand post popularity
- Role of influential users in engaging of a brand post
- Engaging more influential users during the early life of the brand post
- Incentives to influential users
- Release time for products of on the market

Teaching Perspectives

Analytics Ecosystem

Analytics Ecosystem



Data Infrastructure

Category	Examples
Major Hardware and Storage solutions	IBM, Dell, HP, Oracle, EMC, NetApp
Indigenous Hardware and Software Platform	IBM, Oracle, Teradata
Hardware independent Data solutions	Microsoft SQL Server, MySQL
Specialized Integrated Software	SAP
Network Infrastructure/Cloud Computing	Amazon, Salesforce.com
Big Data Infrastructure Services and Training	Cloudera, Hortonworks, Hadoop, Map Reduce, NoSQL

Data Warehouse

Category	Examples
Data integration Services	EMC, IBM, Microsoft, Oracle, SAP, Teradata

Middleware- Business Intelligence Platforms

Category	Examples
Reporting Analytics solutions	IBM Cognos, Microstrategy, Hyperion (Oracle), Plum, SAP Business Objects

Data Aggregators / Distributors

Category	Examples
Data Collection	Comscore, Experian, Google, Nielsen, Omniture,Axiom

Analytics -Application Developers

Category	Examples
Analytics Consulting	IBM, SAS, Teradata
Specific Solutions/Industry Specific	Cerner (Healthcare), IBM Watson (Healthcare and Insurance), Sabre (Travel Industry)
Domain Specific Solutions	Axiom (demographic clustering), FICO & Experian (Credit Score Classification), Demandtec (Pricing Optimization)
Web/Social media/Location Analytics	Sense Networks (location based user profiling), X+1 & Rapleaf (email based profiling), Bluecava (user identification through device usage), Simulmedia (Targeted advertising based on user's TV watching)
Specialized Analytical Solutions	Shazam, Siri(iphone), Google Now(Android)

Analytics User Organizations

Category	Examples
Private Sector, Government, Education, Military etc.	Too numerous to classify

Analytic Industry Analysts and Influencers

Category	Examples
Private Sector, Government, Education, Military etc.	Too numerous to classify
Professional Organizations	Gartner group, Data Warehousing Institute, Forrester, McKinsey
Professional Membership based societies	INFORMS, Special Interest Group on Decision Support Systems (SIGDSS), Teradata, SAS
Analytics Ambassadors	Individual contributors of analytics through seminars, books and other publications

Academic Providers / Certification Agencies

Category	Examples
Analytic Academic Programs	Business school offerings including courses on information systems, marketing, management sciences etc. Computer science, statistics, mathematics, and industrial engineering departmental courses
Analytics Certification programs	IBM, Microsoft, Microstrategy, Oracle, SAS, Teradata

Teaching Perspectives

- Role of universities and academic programs
- Business School offerings in the areas of information systems, Marketing, Management Sciences
- Courses from computer science, statistics, mathematics, and industrial engineering departments
- Academic alliance programs
  - Teradata University Network (teradatauniversitynetwork.com)
- Certificates offered by major vendors

Conclusion

Exciting time for research and teaching in analytics!

Questions/Comments?