- Background and Past Research

- What is Big Data?

- Research Methodology and Findings

- Recommendations

- Department of Defense (DoD) obligated over $240B in FY2015 contracts (USA Spending, 2016)

- USD(AT&L) has called for improving tradecraft in services contracting by strengthening the contracting process

- GAO has identified process deficiencies in DoD documentation and management of CPARS reports
    - Reports are late and are not always completed
    - Report narratives are insufficiently detailed and are, at times, in conflict with associated objective scores

- CPARS deficiencies provide less-than-optimal information to the acquisition team that relies on these reports for source selection and contract administration purposes.
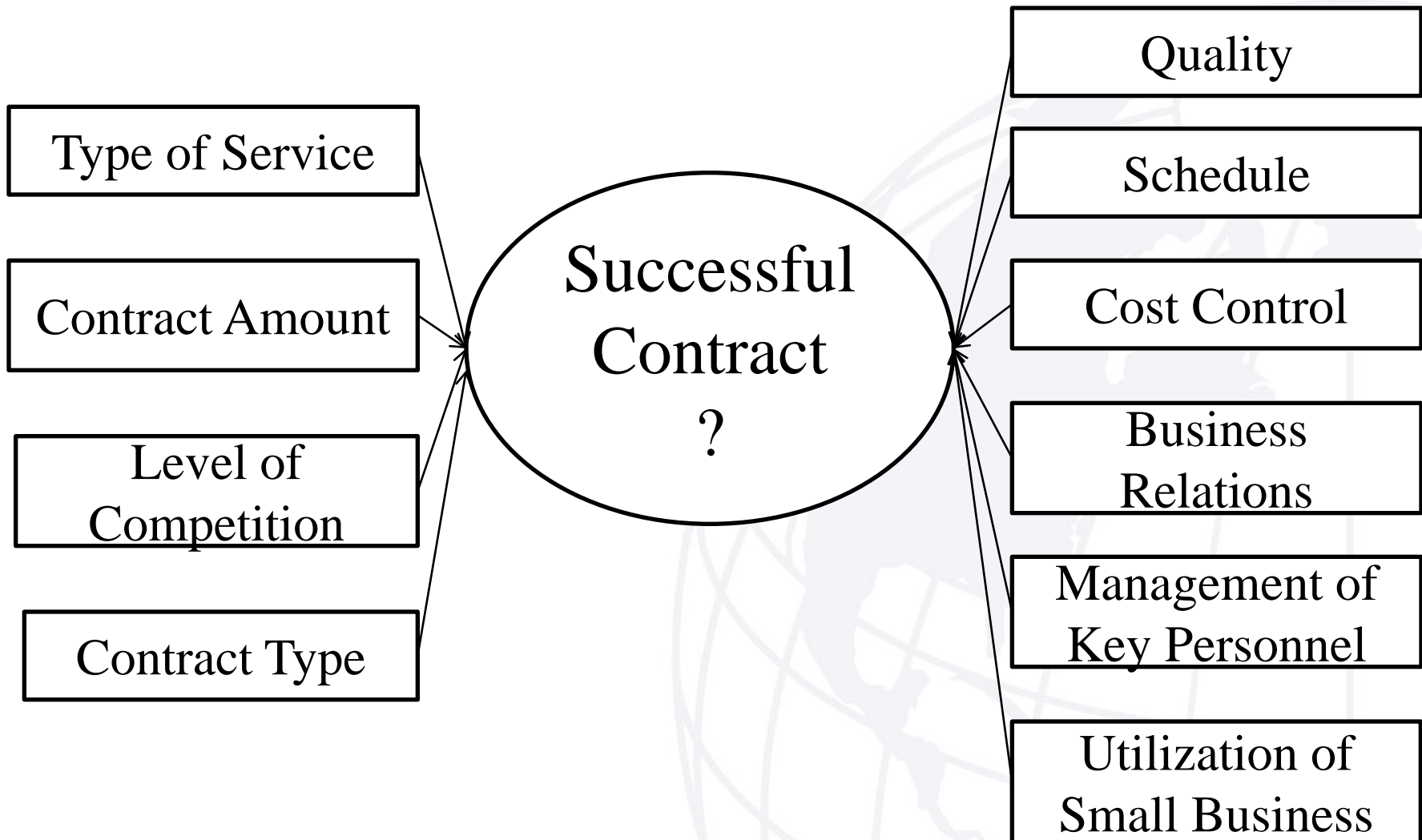
3

- Objective: Identify any relationship between contract variables and contract success.

- Statistical analysis
  - Analyzed 5 MICCs:
    - (Eustis, Knox, Hood, Bragg, and Sam Houston)
  - Analyzed 4 service types:
    - (PAMS; Maintenance/Repair of Equipment; Utilities/Housekeeping; ADP/Telecomm).
  - Analyzed 715 CPARS reports.
  - Investigated the relationship between contract variables and contract success.

Contract Variable

Type of Service

Contract Amount

Level of Competition

Contract Type

Successful Contract ?

CPARS Area

Quality

Schedule

Cost Control

Business Relations

Management of Key Personnel

Utilization of Small Business

- ## Results
  - Utilities/Housekeeping services had the highest failure rate of all the product service codes analyzed.
  - Contracts awarded competitively had the highest failure rate when compared to the other contracts.
  - Contracts structured as a combination contract had the highest failure rate when compared to the other five types of available contracts.
  - As the percentage of 1102 filled billets increased, the contract failure rate decreased.

- ## Limitations
  - Findings based on limited data (only 715 observations)
  - Big Data Analysis techniques is needed for identifying relationships between contract variables and contract success
  - Undertake proof of concept research using Big Data Analysis techniques

Massive influx of data that has been and is currently being collected in the digital and Internet era

90% of the data that is currently being stored on computers and servers around the world was collected in just the <u>past two years</u>

Analytics in a big data world: The essential guide to data science and its applications, Baesens, 2014,

In the <u>year 2000, only one quarter</u> of the world's data was digitized; the remainder was on paper and other analog media. However, <u>by 2013, 98% of all data was digital</u>.

Big data: A revolution that will transform how we live, work, and think. Mayer-Schoenberger & Cukier, 2013

7

The influx of data comes from more:

- digitization,
- interactions,
- communications,
- Internet-consumerism,
- mobile technology,
- social networking.

"Datafication": Turning elements of life into data (pictures, locations, sentiment, etc.)

- Draw inference from <u>large datasets</u> that can be used to:

  1. Make predictions of a "target" variable
  2. Understand relationships between target variables and other "independent" variables

- Large datasets are divided into samples:

  - *Training sample* is used to create an analytical "model"

  - *Validation sample* is used to test the new model

- Multiple "modeling" techniques are used to try to best predict the target variable
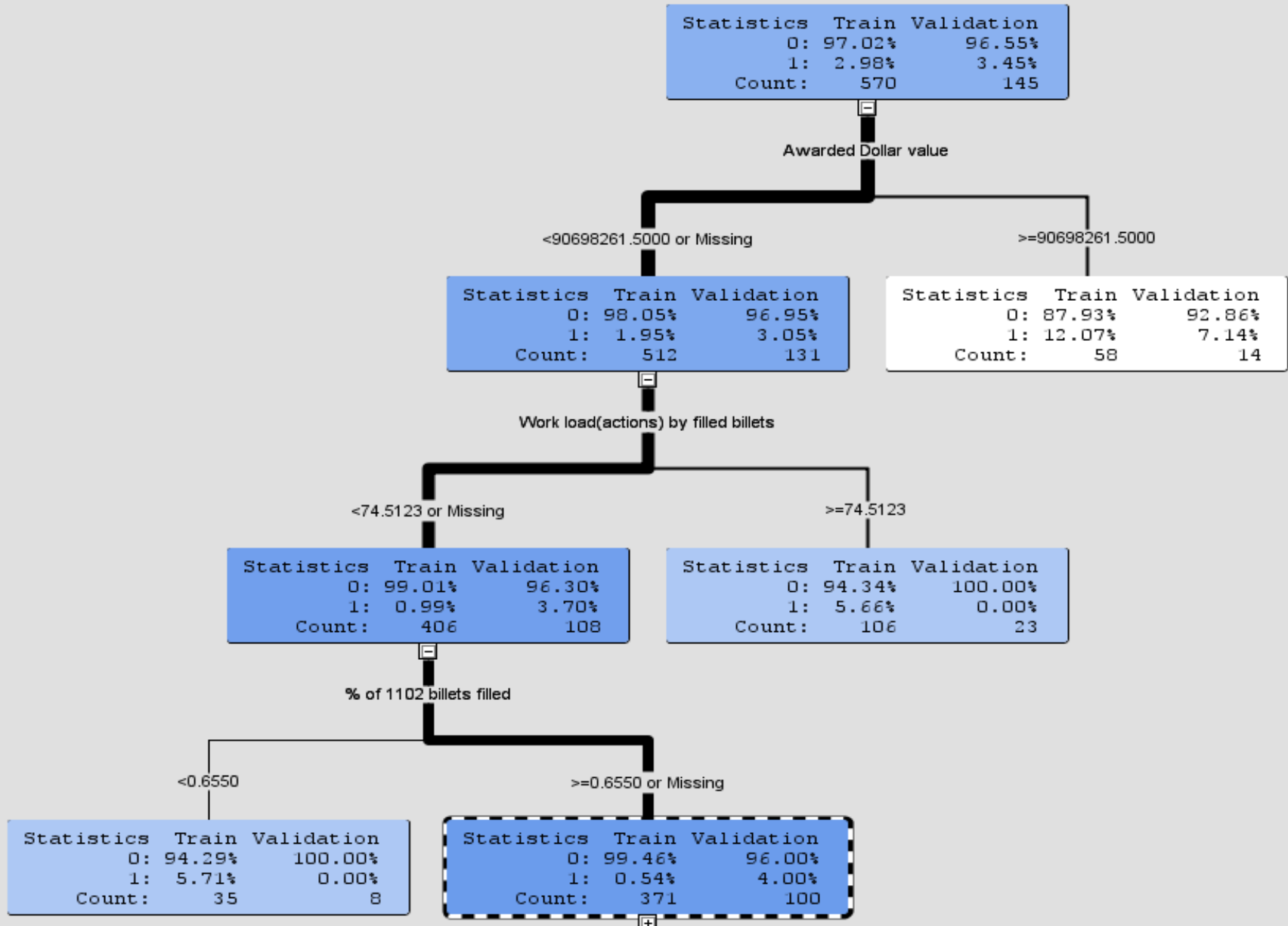
9

- Data description
  - CPAR data combined with MICC data
  - 715 service contracts: 5 MICCS & 4 service codes
  - NOT big data – proof of concept

- Target Variable: Contract Failure

- 20 independent variables

- Modeling Techniques
  - Decision Tree Analysis
  - Logistic Regression
  - Neural Networks

10

- Decision Tree Analysis identifies and isolates groups of observations that act in similar ways in regards to the target variable

- Identifies independent variable that most "discriminate" the target variable

- Divides up the observations into "branches" that further discriminate the target variable along other independent variables.

11

Statistics	Train	Validation
0:	97.02%	96.55%
1:	2.98%	3.45%
Count:	570	145

Awarded Dollar value

<90698261.5000 or Missing

Statistics	Train	Validation
0:	98.05%	96.95%
1:	1.95%	3.05%
Count:	512	131

>=90698261.5000

Statistics	Train	Validation
0:	87.93%	92.86%
1:	12.07%	7.14%
Count:	58	14

Work load(actions) by filled billets

<74.5123 or Missing

Statistics	Train	Validation
0:	99.01%	96.30%
1:	0.99%	3.70%
Count:	406	108

>=74.5123

Statistics	Train	Validation
0:	94.34%	100.00%
1:	5.66%	0.00%
Count:	106	23

% of 1102 billets filled

<0.6550

Statistics	Train	Validation
0:	94.29%	100.00%
1:	5.71%	0.00%
Count:	35	8

>=0.6550 or Missing

Statistics	Train	Validation
0:	99.46%	96.00%
1:	0.54%	4.00%
Count:	371	100

- Linear Regression with a Binomial (0 or 1) target variable.

- Coefficients are interpreted as "odds."

- *Step-wise* methodology runs multiple regression with different independent variable and chooses the one that describes the best with the least variables (parsimony)

13

Results of Logistic Regression

Parameter	Estimate	p value	$e^{(Estimate)}$
Intercept	-12.213	<.0001	0
Work load actions by filled billet	0.0129	0.0117	1.013
Type of Contract – CPAF	8.8507	<.0001	6979
Type of Contract – CPAF & CPFF	-3.2748	0.9986	0.038
Type of Contract – CPFF	9.2498	<.0001	10402
Type of Contract – CPFF FFP	37.0026	0.9954	1.7×10^{16}
Type of Contract – CPIF	-3.3486	0.9978	0.035
Type of Contract – FFP	7.8061	.	2455
Type of Contract - Other	-3.7514	0.9970	0.0264

	Training	Validation
Average Squared Error	0.0266	0.0290
Misclassification Rate	0.0281	0.0276

Series of regression models uncovering latent connecting layers of data that can, in turn, be used to better predict target variables

15
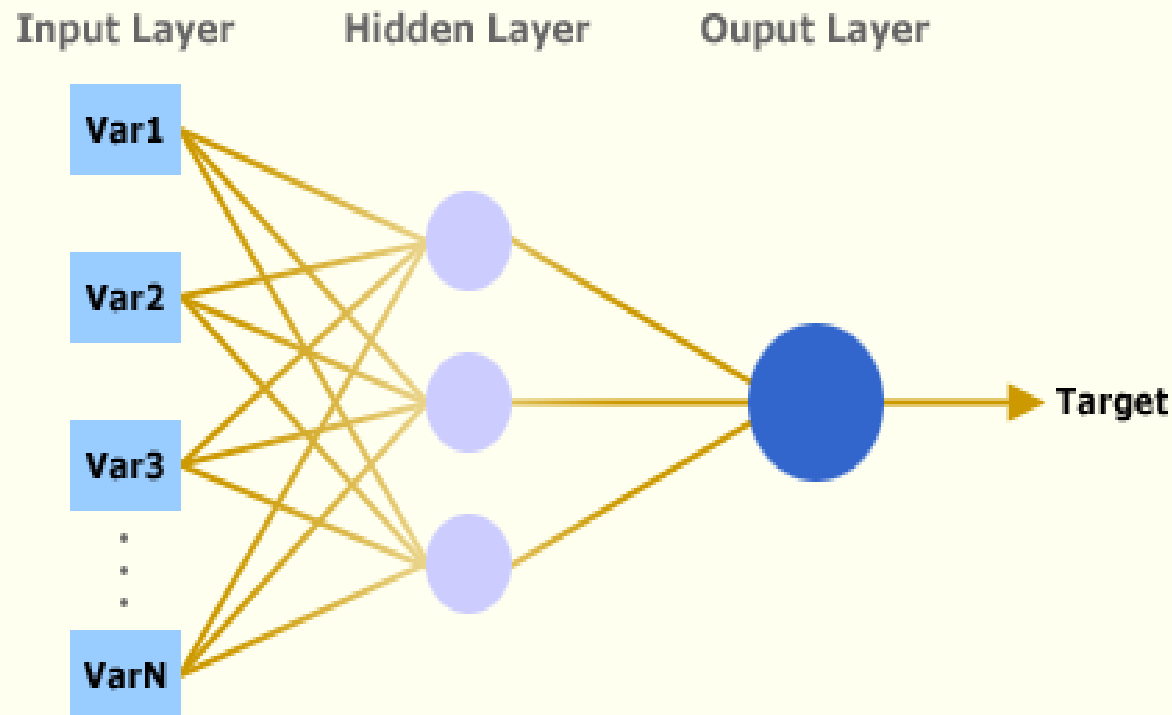
- Due to small data size, the neural network analysis defaulted to the Logistic regression results.

- More data would be needed.

- Proof of Concept show that Big Data analysis could be used for DoD acquisition data.

- Access to databases by researchers is needed. Not just individual CPAR records.

- Other datasets that might be of interest to Big Data Analysis:
  - Source Selection Data (Proposed prices)
  - Selected Acquisition Reports and EVM data
  - FPDS-NG, FEDBIZOPPS,

- Combination of these and other data sources could lead to interesting research questions.

17

Uday M. Apte
Rene G. Rendon

Back Up Slides

	Total Contracts
Total Army MICC Non-System Contracts	14395
Less: Non R, J, S, D Service Contracts	8774
Total R, J, S, D Service Contracts	5621
Less: R, J, S, D Service Contracts at other MICC	4906
R, J, S, D Service Contracts at MICC FDO Eustis, Knox, Hood, Bragg, Sam Houston	715
Fort Eustis	238
Fort Knox	119
Fort Hood	114
Fort Bragg	55
Fort Sam Houston	189

20

- MICC
- Contract Start Month
- Contract Start Day
- Contract Start Year
- Contract End Month
- Contract End Day
- Contract End Year
- Fiscal Year of Contract
- Duration in days
- Contract Type: RJSD
- Awarded Dollar Value

- Current Dollar Value (at time of CPARS)
- Basis of Award
- Type of Contract (FFP, CPFF, CPAF, etc.)
- Annual Workload of Contracting Office (Dollars)
- Annual Workload of Contracting Office (actions)
- # of 1102 Billets Filled by Contracting Office
- % of 1102 Billets Filled by Contracting Office
- Workload ($) by Filled Billet
- Workload (actions) by Filled Billet