

USING DATA ANALYTICS TO DETECT FRAUD

Presentation by:

Denis Yegon Thursday, 12th October,2017

Presentation agenda



Expectations from the session

- ☐ How data analytics can be used to detect fraud
- ☐ Different tools to perform data analytics
- ☐ How to walk through the full data analytics process
- ☐ How to analyze non-numeric data, such as text and timelines, for signs of fraud
- Red flags of fraud that appear in the data
- ☐ Data analytics tests that can be used to detect fraud

Scenario...





True or False



- ☐ Detecting fraud using data analytics is just a simple mouse click away
- ☐ Data analytics saves time
- Data analytics tools can take care of the whole audit process
- ☐ Data analytics replaces the human element in the fraud detection process.



What is Data analytics?



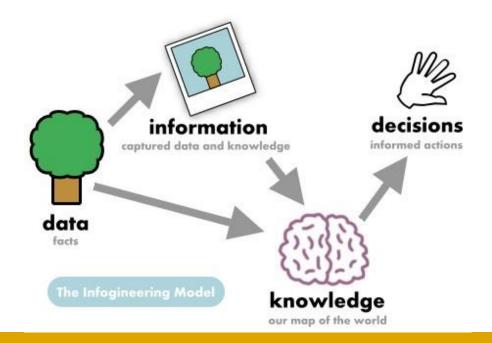
- ☐ Data analytics is the process of examining data in order to formulate conclusions
- ☐ Especially useful when fraud is hidden in large data volumes and manual checks are insufficient
- ☐ Can be used reactively or proactively



What data analytics can do?

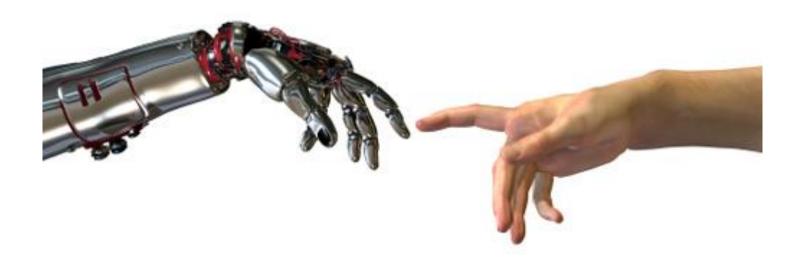


- ☐ Data analytic software enable users to obtain quick overview of data and can drill down to detail information
- Using data analytic software, every transactions can be involved or touched



Scenario....





Expectations....



- ☐ Data analysis techniques alone are unlikely to detect fraud; human judgment is needed to decipher results.
- ☐ Identify anomalies, trends, patterns and concerns
- ☐ Large volumes of data

Types of Data That Can Be Analyzed...



Structured data

- Sales records
- Payment or expense details
- Payroll details
- Inventory records
- Financial reports
- Found in accounting software, databases, spreadsheets, etc.

Unstructured data

- Email and instant messages
- Payment text descriptions
- Social media activity
- Corporate document repositories
- News feeds

Process: Preparing the data....



- ☐ Define the objectives
- Obtain the data
- □ Verify
- Normalize
- □ Analyze
- □ Apply rules/criteria

The process: Finding anomalies....

☐ Report



Resulting anomalies
Interpret the data, professional judgement
Reduce anomalies
Additional criteria
Cut off
Sampling
Investigate
Validate

Standard data analytic features...



- ☐ Extract
- ☐ Sort
- ☐ Gaps
- Duplicates
- ☐ Aging
- ☐ Samples
- Summarize
- ☐ Stratify
- ☐ Join/Match

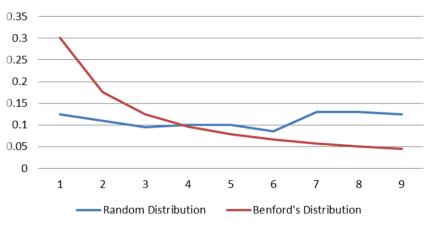


Statistical & advanced features...



- ☐ Benford's Law
 - ☐ Most naturally occurring data sets follow a strange rule called Benford's Law
- ☐ Trend Analysis
- ☐ Time Series Analysis
- Correlation





Artificial Intelligence and statistical modeling



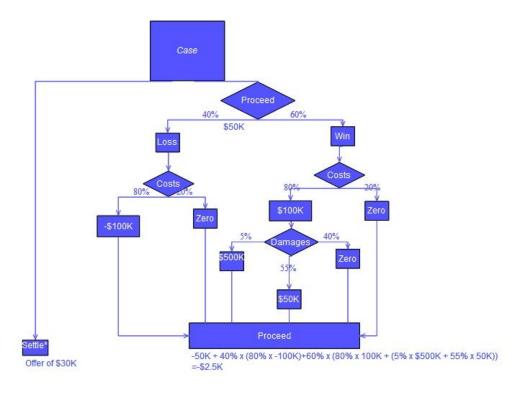
- ☐ Supervised Machine Learning
- ☐ Unsupervised Machine learning
- ☐ Regression analysis



Supervised machine learning



- ☐ Supervised Machine Learning
- ☐ Example, decision trees



Unsupervised machine learning



- ☐ The machine is not given any rules or instructions to guide it's learning. Machine learns through "experience".
- ☐ Examples include deep learning.
- Growing at an unprecedented rate...



Mankind's biggest threat?





- ☐ Space X
- ☐ Tesla
- ☐ Solar City
- ☐ Open AI





Worth reading Superintelligence by Bostrom. We need to be super careful with AI. Potentially more dangerous than nukes.



Scenario....



Employee			Direct Deposit	Deposit
ID	Employee Name	Loction	Account	A mount
10078	William Kayak	Buffalo, NY	000756227	3,243
11265	Edward Cook	Miami, FL	000756227	5,538
13655	Nancy Wright	Chicago, IL	000756227	2,236

Total Paid: 11,017

Data analytic tools

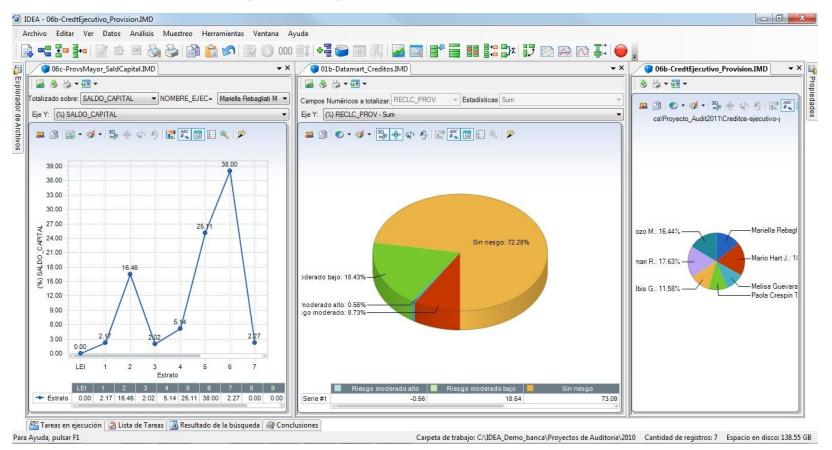


- ☐ These include;
- ☐ CAAT: IDEA, ACL, Arbutus
- ☐ Analytics: Alteryx, Rapidminer, SAS, Knime, IBM
- ☐ Visualization: Tablue, Qlik, PowerBi
- ☐ Programing Languages: R, Python

Data analytic tools



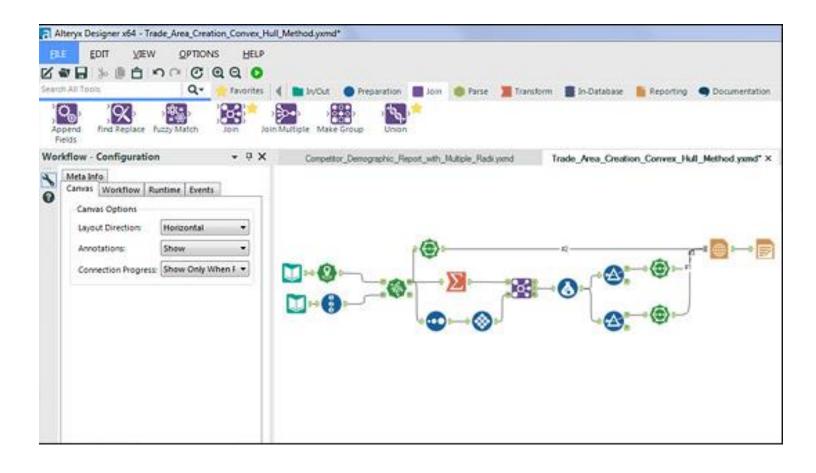
☐ CAAT: IDEA, ACL, Arbutus



Analytics: Alteryx, Rapidminer, SAS, Knime, IBM



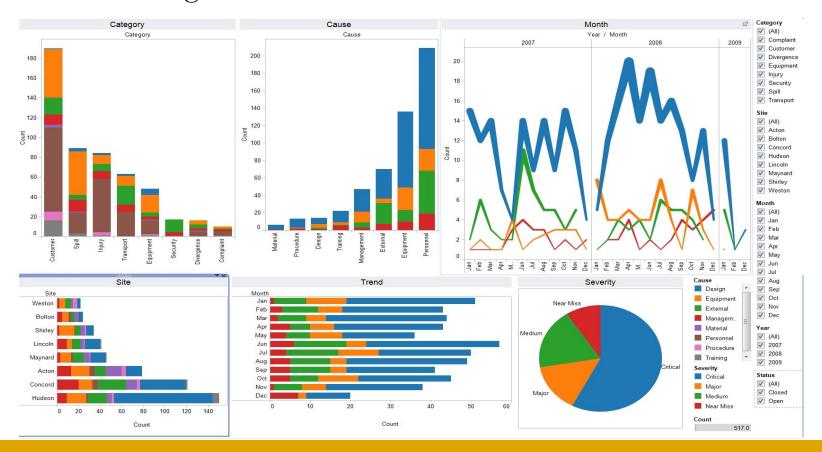
☐ Analytics: Alteryx, Rapidminer, SAS, Knime, IBM



Visualization: Tablue, Qlik, Power



☐ Can be used for creation of dashboards for active monitoring.



Programing languages: R, Python



☐ Working with large datasets, or creating complex data visualizations, they are godsends.

```
Returns a QuerySet of connections for user.

set1 = self.filter(from_user=user).select_related(depth=1)

set2 = self.filter(to_user=user).select_related(depth=1)

def are_connected(self, user1, user2):
    return True
    if self.filter(from_user=user1, to_user=user2).count() > 0:
        return True
    if self.filter(from_user=user2, to_user=user1).count() > 0:

    def remove(self, user1, user2):

    Deletes proper object regardless of the order of users in one if not connection:
```

Scenario...





Challenges.



- ☐ "Dirty" data
- ☐ Skills gap
- ☐ Security and privacy
- ☐ Internal silos

Solutions.



- ☐ Better data governance
- ☐ Training and awareness
- ☐ Stringent security and privacy controls
- ☐ Improved internal collaboration

Summary



- ☐ By using anomaly detection methods, data analytics can help in finding fraud
- Unusual patterns needs to be identified
- ☐ With fraudulent transactions being rare, one must understand what is normal

