<table>
<thead>
<tr>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>Background and experience.</td>
</tr>
<tr>
<td><strong>What is Data Mining?</strong></td>
</tr>
<tr>
<td>High level overview.</td>
</tr>
<tr>
<td><strong>Decision Tree Examples</strong></td>
</tr>
<tr>
<td>Overview of the decision tree technique, and how I have applied it in different ways.</td>
</tr>
<tr>
<td><strong>Final Thoughts</strong></td>
</tr>
<tr>
<td>Lessons learned.</td>
</tr>
</tbody>
</table>
Agenda

Introduction
Background and experience.

What is Data Mining?
High level overview.

Decision Tree Examples
Overview of the decision tree technique, and how I have applied it in different ways.

Final Thoughts
Lessons learned.
Introduction

Jason Noriega

Senior Workforce Analyst at SanDisk

Experience: eBay, Lawrence Livermore National Laboratory, NASA, and National Geospatial Intelligence Agency.
Agenda

Introduction
Background and experience.

What is Data Mining?
High level overview.

Decision Tree Examples
Overview of the decision tree technique, and how I have applied it in different ways.

Final Thoughts
Lessons learned.
What is Data Mining?

Data mining is a field that utilizes methods of machine learning, traditional statistics, database technology, and data visualization.

Typical Uses:
- Targeted Customer Retention
- Predict 12 Months of Sales of a New Product
- Building Effective Marketing Campaigns

WF Analytics:
- Targeted Employee Retention
- Predict 12 Months of Employee Separations
- Building Effective Recruiting Initiatives
High Level Process

Historical Data

• R Program
• Weka
• SAS EG
• SAS JMP
• SPSS Modeler

Predictive Model

Data with Predictions
Agenda

Introduction
Background and experience.

What is Data Mining?
High level overview.

Decision Tree Examples
Overview of the decision tree technique, and how I have applied it in different ways.

Final Thoughts
Lessons learned.
Decision Tree Technique

Decision tree algorithms applied to employee attrition data provides you with a **laser precision focus** on employee retention.

**Recursive Partitioning**

**Predictors**
1. Job Function
2. Gender
3. Ranking
4. ..... 

**Predictors**
1. Age
2. Rating
3. Commute
4. ..... 

**Predictors**
1. Performance
2. Tenure
3. Commute

**Decision Rules - Example**

**Rule 1: 50% Likelihood of Leaving**
If Job Function = **Software Engineer**
And Age < 40
And Performance = **Top Performer**

**Rule 2: 60% Likelihood of Leaving**
If Manager Effectiveness < 70% Favorable
And Pay Grade < 17, And Region = **Americas**

**Rule 3: 70% Likelihood of Leaving**
If Not Promoted in Prior 2 Years = **No**
And Tenure > 3 Years

All data is notional
Improving the Retention of Employees

Example 1

Decision Tree Analysis
Where are the high rates of voluntary turnover?

Exit Survey
Why do these patterns exist?

Prioritization
Where should we prioritize areas of improvement?

Action Planning
What can do to reduce high rates of voluntary turnover?
Example 1 - Variables

Variables for Attrition Analysis

Demographics/Tenure
- Age
- Gender
- Tenure

Job Content
- Reasgned Diff. Pers Area
- Job Function
- Exempt/Nonexempt
- Latest Perform. Rating
- Perform. Rating Delta
- Manager Flag

Location
- Region
- Country
- Location

Compensation
- Pay Grade
- Comp Ratio
- Promotion Flag
- Spot Award Flag
- Tuition Reimbursement Flag

Manager
- Mgr Tenure
- Mgr Perform.

Attrition Analysis Timeframe

- 2010
- 2011
- 2012

Model Timeframe for Finding Patterns

May
June
Example 1 - Tree

Employee Tenure

- Less than 1.1 yrs
  - #           %
  - Stayed: 4000 80%
  - Left: 1000 20%

- 1.1-2.6
  - #           %
  - Stayed: 3000 86%
  - Left: 500 14%

- 2.6-7.3
  - #           %
  - Stayed: 3000 94%
  - Left: 250 8%

- > 7.3 Yrs
  - #           %
  - Stayed: 1000 95%
  - Left: 50 5%

Pay Grade

- <= 16
  - #           %
  - Stayed: 1500 71%
  - Left: 600 29%

- 17-22
  - #           %
  - Stayed: 500 82%
  - Left: 100 17%

- > 22
  - #           %
  - Stayed: 2000 87%
  - Left: 300 13%

Business Unit

- Organization #1
  - #           %
  - Stayed: 1100 85%
  - Left: 200 15%

- Organization #2
  - #           %
  - Stayed: 400 50%
  - Left: 400 50%

Compa Ratio

- <= 85
  - #           %
  - Stayed: 100 33%
  - Left: 200 67%

- > 85
  - #           %
  - Stayed: 300 75%
  - Left: 100 25%

All data is notional
Example 1 – Survey Linkage

**Employee Tenure**
- **Less than 1.1 yrs**
  - Stayed: 4000 80%
  - Left: 1000 20%
- **1.1-2.6 yrs**
- **2.6-7.3 yrs**
- **> 7.3 Yrs**
  - Stayed: 1000 95%
  - Left: 50 5%

**Pay Grade**
- **<= 16**
  - Stayed: 1500 71%
  - Left: 600 29%
- **17-22**
  - Stayed: 500 82%
  - Left: 100 17%
- **> 22**
  - Stayed: 2000 87%
  - Left: 300 13%

**Exiting Employees**
- Low Importance/High Performance
- High Importance/High Performance
- Low Importance/Low Performance
- High Importance/Low Performance
- Uncompetitive Pay
- Job matched expectations
- Opportunity for growth

All data is notional
Example 1 - Tree

Employee Tenure

- Less than 1.1 yrs
  - Stayed: 4000 80%
  - Left: 1000 20%

- 1.1-2.6 yrs
  - Stayed: 3000 86%
  - Left: 500 14%

- 2.6-7.3 yrs
  - Stayed: 3000 90%
  - Left: 250 10%

- > 7.3 Yrs
  - Stayed: 1000 93%
  - Left: 50 7%

Performance Rating

- DNM, Meets Some
  - Stayed: 300 67%
  - Left: 150 33%

- Meets
  - Stayed: 1500 82%
  - Left: 200 13%

- Exceeds Some
  - Stayed: 1500 82%
  - Left: 200 13%

- Exceeds Most
  - Stayed: 1200 93%
  - Left: 150 7%

Promotion

- Yes
  - Stayed: 850 95%
  - Left: 50 5%

- No
  - Stayed: 350 78%
  - Left: 100 22%

Big Data Forbes Article
Example 2 – Grouping Variables

<table>
<thead>
<tr>
<th>Location/Job/Org</th>
<th>Manager Impact</th>
<th>Recruitment</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Rating</td>
<td><strong>Overall Engagement Impact</strong></td>
<td>Education Level</td>
<td>Age</td>
</tr>
<tr>
<td>Region</td>
<td>- Overall Satisfaction</td>
<td>Total Yrs Experience</td>
<td>Gender</td>
</tr>
<tr>
<td>State</td>
<td>- Feel Proud</td>
<td>Total Yrs Relevant Experience</td>
<td>Compa Ratio</td>
</tr>
<tr>
<td>Country</td>
<td>- Feel Motivated</td>
<td>Recruitment Source</td>
<td></td>
</tr>
<tr>
<td>Work Location</td>
<td>- Recommend as Great Place</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pay Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Function</td>
<td><strong>Overall Job Sat. Score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Level</td>
<td>- Work feels important purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Unit</td>
<td>- Clearly understand indv. Objectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Job makes good use of skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Opportunity to learn and grow</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Overall Mgr Sat. Score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mgr clearly explains expectations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mgr provides ongoing coaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mgr actively supports prof. dev.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mgr involve people with decisions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mgr instills positive attitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mgr is someone I trust</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Mgr Performance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Mgr Tenure</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where does the problem exist?

What impact does the mgr have?

How can we target those most likely to stay?

Are there any other important variables?

Impact of People Mgr

Recruitment

Other

Location, Job, Org, Other

SanDisk

All data is notional
Where Does the Problem Exist?

**Location/Job/Org**
- Perform. Rating
- Region
- State
- Country
- Work Location
- Pay Grade
- Job Function
- Job Level
- Business Unit

**Performance Rating**

- **Does Not Meet**
  - # | %
  - Stayed: 84 76%
  - Left: 26 24%

- **Meets Some, Meets, Exceeds Some/Most**
  - # | %
  - Stayed: 1210 83%
  - Left: 252 17%

**Region**

- **Americas**
  - # | %
  - Stayed: 795 80%
  - Left: 194 20%

- **Asia Pacific**
  - # | %
  - Stayed: 83 87%
  - Left: 12 13%

- **EMEA**
  - # | %
  - Stayed: 248 93%
  - Left: 20 7%

**Job Function**

- **A, B, and C**
  - # | %
  - Stayed: 638 77%
  - Left: 188 23%

- **All Other**
  - # | %
  - Stayed: 157 96%
  - Left: 6 4%
What Impact does the Manager Have?

**Mgr Impact**
- Overall Engagement Score
  - Overall Satisfaction
  - Feeling Proud
  - Feeling Motivated
  - Recommend great place

- Overall Job Sat. Score
  - Work feels important purpose
  - Clearly understand indv. Obj
  - Job makes good use of skills
  - Oppty to learn and grow

- Overall Mgr Effectiveness
  - Mgr clearly explains expectations
  - Mgr provides ongoing coaching
  - Mgr actively supports prof. dev.
  - Mgr involve people with decs
  - Mgr instills positive attitude
  - Mgr is someone I trust

- Mgr Performance
- Mgr Tenure

---

**Performance Rating**

- Does Not Meet
  - # %
  - Stayed: 84 76%
  - Left: 26 24%

- Meets Some, Meets, Exceeds Some/Most
  - # %
  - Stayed: 1126 83%
  - Left: 226 17%

**Region**

- Americas
  - # %
  - Stayed: 795 80%
  - Left: 194 20%

- Asia Pacific
  - # %
  - Stayed: 83 87%
  - Left: 12 13%

- EMEA
  - # %
  - Stayed: 248 93%
  - Left: 20 7%

**Job Function**

- A, B, and C
  - # %
  - Stayed: 638 77%
  - Left: 188 23%

- All Other
  - # %
  - Stayed: 157 96%
  - Left: 6 4%

**Mgr Effectiveness**

- <= 75% Favorable
  - # %
  - Stayed: 351 73%
  - Left: 128 27%

- > 75% Favorable
  - # %
  - Stayed: 287 84%
  - Left: 60 16%

---

All data is notional
Example 3  Quarterly Workforce Dashboard with Deep Dive

Quarterly Dashboard

Deep Dive

Predictors
1. Tenure
2. Age
3. Region
4. ...

Predictors
1. Region
2. Rating
3. Pay Grade
4. ...

Predictors
1. Grade
2. Performance
3. Age

Stayers 81%
Leavers 19%

<= 2.5 yrs  Tenure  > 2.5 yrs

Stayers 75%
Leavers 25%

<= 7  Grade  > 7

Stayers 65%
Leavers 35%

<2013-Q2 2013-Q3 2013-Q4 2014-Q1 2014-Q2

Voluntary-ORG Involuntary-ORG Voluntary-Companywide

All data is notional
Software

R Program for Statistical Computing

http://www.r-project.org/
http://rattle.togaware.com/rattle-install-mswindows.html

SPSS Modeler