Data Science and Technology Entrepreneurship
(Spring 2013)

Business Model, Customer Validation, Linear Classifiers, Intellectual Property

Sameer Maskey
Week 2
Announcement

- Looking for the 3rd TA
  - Computer Science
  - Experience in Machine Learning, MapReduce, Web programming
- If you have friends who may be interested let them know
Topics for Today

‣ Guest Lecture
  ‣ Intellectual Property, Patents, Columbia Technology Ventures

‣ Business Model

‣ Customer Validation

‣ Linear Classifiers
Guest Lecture

- Orin Herskowitz
- Executive Director, Columbia Technology Ventures (CTV)
Assignment 1 - Due Friday

- Assignment 1 was posted on the course website
- Due this Friday @6pm

Submission Method

- Each team create a google id
  - teamXYZ@gmail.com
- put your doc in teams google drive
- send us a shared link (email Sameer, Morgan and Jigar)

Email Subject Line:

- DSTE - Assignment 1 - TeamName
Extra Classes: Web Programming 101

- Starts in 1 or 2 weeks
- 2-3 extra lectures
- 1-2 hour lecture each
- Fridays @3:30 pm

Web Programming Basics

- If you know how to build web applications you don’t have to come
- Geared for MBA students who want to know bit more behind technology that runs new web startups
- Geared for CS/Eng students who know programming but not necessarily web programming
Teams

- Minimum (1 MBA + 1 CS)
- Maximum (2 MBA + 2 CS) - Not a strict rule
- Team of 4+?
  - Let me know
Team Name

- As a part of your assignment you are supposed to come with a name for your team/start up
  - Test the name with your friends and family
    - Right name?
  - domain available?
  - social handles - facebook, twitter, handles available?
Reminder: Textbook

- Technology Ventures: From Idea to Enterprise, 3rd edition
- Thomas Byers (Author), Richard Dorf (Author), Andrew Nelson (Author)
Course Stages

**Stage 1** (3 weeks – Jan 30 – March Feb 20) Problem definition, Data collection, Customer development, Business Model Canvas, Data science methods for testing your hypothesis

**Stage 2** (5 weeks – Feb 4 – March 10) Minimum Viable Product development, Quantifying customer feedback with classification and clustering techniques

**Stage 3** (2 weeks – March 11 – March 31) Agile development, Data analysis of feature surveys, Sequential prediction algorithms (costs, revenue, traction)

**Stage 4** (2 weeks – April 1 – April 29) Launching the product, Data driven marketing techniques, A/B testing

**Stage 5** (2 weeks – April 1 – May 5) Try to raise funds with VC network provided in the class
Customer/Market Risk vs Invention Risk

<table>
<thead>
<tr>
<th>Web 2.0</th>
<th>Enterprise Software</th>
<th>Enterprise Hardware</th>
<th>Comm Hardware</th>
<th>Comm Software</th>
<th>Consumer Electronics</th>
<th>Game Software</th>
<th>Semicon</th>
<th>Electronic Design Automation</th>
<th>Cleantech</th>
<th>Med Dev / Health Care</th>
<th>Life Science Biotech</th>
</tr>
</thead>
</table>

Where does your startup fall?

Source: Steve Blank
Customer/Market Risk vs. Invention Risk

More Invention Risk you have, more important Intellectual Property become
Guest Lecture

- Orin Herskowitz’s slides are posted separately in the course website
Business Model and Startups

- When you try to build a startup you can go through a business design exercise [Byers, et. al]
  - who are the customers?
  - how are needs of customers satisfied?
  - how are profits captured and profitability protected?
- What is a business model?
Business Model and Startups

‣ When you try to build a startup you can go through a business design exercise [Byers, et. al]
  ➢ who are the customers?
  ➢ how are needs of customers satisfied?
  ➢ how are profits captured and profitability protected?

‣ What is a business model?

“A business model is a set of planned assumptions about how a firm will create value for all its stakeholders.” [Magretta 2002]
Business Model Hypotheses

- When you have a startup you have assumptions
Business Model Hypotheses

- When you have a startup you have assumptions

  - My users will share funny pictures at least 20 times
  - My value to the client is the speed of our system
  - Our recommendation engine works well

  -
  -
  -

  - Our revenue will hit 200K US$ in 4 months
Business Model Hypotheses

- When you have a startup you have assumptions

My users will share funny pictures at least 20 times
My value to the client is the speed of our system
Our recommendation engine works well

Our revenue will hit 200K US$ in 4 months

These assumptions need to be validated
In fact, in the beginning, many assumptions tend to be off the mark or outright wrong
Business Models - Variation in View

- Different interpretations on what constitutes a business model
- What assumptions are relevant?
- We are going to focus on ones that are particularly relevant for startups
  - Elements of Business Model [Byers, et. al]
  - Business Model Canvas [Osterwalder, A]
  - Lean Canvas [Maurya, A]
<table>
<thead>
<tr>
<th>Element</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer selection</td>
<td>Who is the customer? Is our offering relevant to this customer?</td>
</tr>
<tr>
<td>Value proposition</td>
<td>What are the unique benefits?</td>
</tr>
<tr>
<td>Differentiation and control</td>
<td>How do we protect our cash flow and relationships? Do we have a sustainable competitive advantage?</td>
</tr>
<tr>
<td>Scope of product and activities</td>
<td>What is the scope of our product activities? What activities do we do, and what do we outsource?</td>
</tr>
<tr>
<td>Organizational design</td>
<td>What is the organizational architecture of the firm?</td>
</tr>
<tr>
<td>Value capture for profit</td>
<td>How does the firm capture some of the total value for profit? How does the firm protect this profitability?</td>
</tr>
<tr>
<td>Value for talent</td>
<td>Why will good people choose to work here? How will we leverage their talent?</td>
</tr>
</tbody>
</table>
## Business Model Canvas [Osterwalder, A]

<table>
<thead>
<tr>
<th>Key Partnerships</th>
<th>Key Activities</th>
<th>Value Propositions</th>
<th>Customer Relationships</th>
<th>Customer Segments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key Resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cost Structures</th>
<th>Revenue Streams</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Image Source: Web [FiveWhys]
Lean Canvas [Maurya, A]

<table>
<thead>
<tr>
<th>PROBLEM</th>
<th>SOLUTION</th>
<th>UNIQUE VALUE PROPOSITION</th>
<th>UNFAIR ADVANTAGE</th>
<th>CUSTOMER SEGMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 3 problems</td>
<td>Top 3 features</td>
<td>Single, clear, compelling message that states why you are different and worth buying</td>
<td>Can’t be easily copied or bought</td>
<td>Target customers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEY METRICS</th>
<th>CHANNELS</th>
<th>COST STRUCTURE</th>
<th>REVENUE STREAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key activities you measure</td>
<td>Path to customers</td>
<td>Customer Acquisition Costs</td>
<td>Revenue Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Distributing Costs</td>
<td>Lifetime Value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hosting</td>
<td>Revenue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>People, etc.</td>
<td>Gross Margin</td>
</tr>
</tbody>
</table>

Lean Canvas is adapted from The Business Model Canvas (http://www.businessmodelgeneration.com) and is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported License.
Common Concept Across Frameworks

- Differentiations and Control
- Product Activities
- Customer Selection
- Organizational Design
- Value Proposition
- Value for talent
- Value Capture for Profit

Elements of Business Model

- Value Proposition
- Customer Segments
- Key Partners
- Revenue Streams
- Key Activities
- Customer Relationship
- Key Resources
- Cost Structure
- Channels
- Unfair Advantage
- Key Metrics
- Revenue Streams

Lean Canvas

Business Model Canvas
Common Concept Across Frameworks

- Value for Talent
- Organizational Design
- Value Capture for Profit
- Key Activities
- Unfair Advantage
- Customer Proposition
- Customer Segments
- Key Metrics
- Problem
- Solution
- Key Partners
- Customer Relationship
- Key Resources
- Channels
- Revenue Streams
- Cost Structure
Value Proposition & Customer Segments

- This is a common theme across all business model frameworks.

- 2 important questions to ask yourself:
  - Who are your customers?
  - What is the value proposition for them?
Customers

Customer Segments

- Who are your customers?
- Mass market? Niche?
- People, Small Businesses (SMEs), Large Corporations?
  - Business travelers?
  - SMEs in retail industry?
    - Food?
      - Canned food?
      - Beans?
      - Organic?
      - ...

- Customer segments can be dissected or categorized in many ways
- How to really find out your assumptions about customer segment are correct?
Value Proposition

- Next relevant question is: Value Proposition for your customer?
- What features of your product provides unique value to your customer?
- Why should they use or buy your product?
  - Lowest price?
  - Make things easier?
  - Fastest?
  - Newest technology?
  - Customization?
Value Proposition

- Founder Ken Croc: “McDonalds stands for friendliness, cleanliness, consistency, and convenience”
- Coca-Cola - unique taste
- Ryanair - extremely cheap
- Walmart - price
- Your company’s value proposition?
## Five Key Values

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Product</td>
<td>Performance, Quality, Features, Brand, Selection, Search, Safe, Ease</td>
</tr>
<tr>
<td>2</td>
<td>Price</td>
<td>Fair, Visible, Consistent, Reasonable</td>
</tr>
<tr>
<td>3</td>
<td>Access</td>
<td>Convenient, Location, Nearby, At-hand, Easy to find, in a reasonable time</td>
</tr>
<tr>
<td>4</td>
<td>Service</td>
<td>Ordering, Delivery, Return, Check-out</td>
</tr>
<tr>
<td>5</td>
<td>Experience</td>
<td>Emotional, Respect, Ambiance, Fun, Intimacy, Relationships, Community</td>
</tr>
</tbody>
</table>

Source: Tech Ventures book
## Primary and Secondary Value

<table>
<thead>
<tr>
<th>Secondary Value</th>
<th>Product</th>
<th>Price</th>
<th>Access</th>
<th>Service</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Target</td>
<td>Wal-Mart</td>
<td>Amazon.com</td>
<td>Honda</td>
<td>Harley-Davidson</td>
</tr>
<tr>
<td>Access</td>
<td>Google</td>
<td>Priceline</td>
<td>Holiday Inn</td>
<td>Wal-Mart</td>
<td>Disney World</td>
</tr>
<tr>
<td>Access</td>
<td>Barnes &amp; Noble</td>
<td>Visa</td>
<td></td>
<td>Dell Computer</td>
<td>Olive Garden</td>
</tr>
<tr>
<td>Service</td>
<td>Toyota</td>
<td></td>
<td></td>
<td></td>
<td>Starbucks</td>
</tr>
<tr>
<td>Service</td>
<td>Home Depot</td>
<td>Southwest Airlines</td>
<td>McDonald’s</td>
<td></td>
<td>Carnival Cruise Line</td>
</tr>
<tr>
<td>Experience</td>
<td>Mercedes</td>
<td>Virgin Atlantic</td>
<td>AT&amp;T</td>
<td>Nordstrom</td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>Best Buy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Byers et. al. book
Articulating Value Proposition

Our company, (company name), is developing (a defined offering) to help (a target audience) (solve a problem) (with secret sauce)

Adeo Ressi’s Mad Lib Template
Lean Canvas [Maurya, A]

https://docs.google.com/drawings/d/1RCczinVGbE1Fj0geyOwpGWWM5FYkvmLSXnRenf9dY_o/edit
Are Your Assumptions Valid?

- Assume you have come up with a great value proposition
- How can you validate that value proposition is really valid?
  - Validate by testing your hypotheses with customers!

For more read: Startup Owner’s Manual
Customer Discovery Process

- Source: Startup Owner’s Manual - Steve Blank and Bob Dorf

Can we use data driven methods? What kind of methods?
Some startups are already data driven
Recap: Data to Scores

- We looked at how we can use a linear regression model to predict if a product is likely to clicked.
- Converted raw data (product information) into insights (likelihood of clicks).
- The method of scoring raw data can also be used to test if your customer acquisition rate prediction is valid.
- To test if your pricing model is valid.
Data to Scores

- Raw Data => Processed Data => Model => Prediction Score

- Given our training data
  - (1, 4)
  - (0, 2)
  - .
  - .
  - .
  - (1, 9)

Training Our Regression Model:
Just need to implement for loop that computes numerators and denominators in equations here. And we get optimal thetas

\[
\theta_1 = \frac{\sum_{i=1}^{N} x_i y_i - \frac{1}{N} \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i^2 - \frac{1}{N} \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i}
\]

For Prediction/Testing:
Given optimal thetas, plug in the x value in our equation to get y

\[
\theta_0 = \frac{1}{N} \sum_{i=1}^{N} y_i - \frac{1}{N} \theta_1 \sum_{i=1}^{N} x_i
\]
## User Data History

<table>
<thead>
<tr>
<th>Product</th>
<th>Clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stauer Compendium Hybrid Watch</td>
<td>4</td>
</tr>
<tr>
<td>Michael Kors Watches Bradshaw Gold - Michael</td>
<td>5</td>
</tr>
<tr>
<td>Men's Pro Diver Chronograph Black</td>
<td>2</td>
</tr>
</tbody>
</table>
Process Raw Data

Raw Data

Product, Clicks

(Stauer Compendium Hybrid Watch, 4)

(Michael Kors Watches Bradshaw Gold - Michael, 5)

(Men's Pro Diver Chronograph Black, 2)

Processed Data

(x_i, y_i)

Product, Clicks

(1, 4)

(1, 5)

(0, 2)

Process

Count number of word “Watch” in Product Description
Data to Predicted Scores

Raw Data

\[(?,?)\]

Processed Data

\[(?,?)\]

\[f(x) = \theta_0 + \theta_1 x\]

\[\theta_0 = 0; \theta_1 = 2\]

Our Model

This watch will be potentially clicked twice

\[\text{This watch will be potentially clicked twice}\]
Business Model Assumptions

- The company Zoolaster sells Zoola watches online.
- Zoolaster buys watches from wholesaler for cheaper price and sells them online.
- Zoolaster assumes certain types of Zoola watches sell well.
- Zoolaster executives want to quantify which Zoola watches may sell well so that they just buy those kind from the wholesaler.
Sales Data

Product | Buy?
---------|-----
Stauer Compendium Hybrid Watch | 1
Michael Kors Watches Bradshaw Gold - Michael | 1
Men's Pro Diver Chronograph Black | 0

1 = Bought
0 = Didn’t Buy
Zoolaster can potentially buy more watches from wholesaler that have higher potential of selling online.
Machine Learning and Business Model Testing

- We will look at four main ways to analyze data that are all useful in decision making for businesses in general:
  - Data to Scores
  - Data to Classes
    - Discriminative Methods
    - Generative Methods
  - Data to Clusters
Data to Classification

- Given a set of features
  \[ X = (x_1, x_2, x_3, \ldots, x_n) \]

- we want to predict \( Y \)
  \[ Y = \{0, 1\} \]

How about \( x \)?
How do we get them?
Data to Classification

- Given a set of features

\[ X = (x_1, x_2, x_3, \ldots, x_n) \]

- We want to predict Y

\[ Y = \{0, 1\} \]

How about x? How do we get them?

\( \{ \text{will not buy OR will buy} \} \)
Can we build a regression model to model such binary classes?

Train Regression and threshold the output

- If $f(x) \geq 0.7$ CLASS1
- If $f(x) < 0.7$ CLASS2
- $f(x) \geq 0.5$ ?
Data to Classification

- Can we build a regression model to model such binary classes?
- Train Regression and threshold the output
  - If $f(x) \geq 0.7$ CLASS1
  - If $f(x) < 0.7$ CLASS2
  - $f(x) \geq 0.5$ ?

\[ \begin{align*}
\text{Happy/Good/ClassA} & \quad f(x) \geq 0.5? \\
\text{Sad/Not Good/ClassB} &
\end{align*} \]
Generative vs. Discriminative Classifier
Generative vs. Discriminative Classifier

- Generative Classifier
Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability $p(x,y)$ where $x$ are inputs and $y$ are labels
Generative vs. Discriminative Classifier

- **Generative Classifier**
  - Model joint probability $p(x,y)$ where $x$ are inputs and $y$ are labels
  - Make prediction using Bayes rule to compute $p(y|x)$
Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability $p(x,y)$ where $x$ are inputs and $y$ are labels
  - Make prediction using Bayes rule to compute $p(y|x)$

- Discriminative Classifier
Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability $p(x,y)$ where $x$ are inputs and $y$ are labels
  - Make prediction using Bayes rule to compute $p(y|x)$

- Discriminative Classifier
  - Try to predict output directly
Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability $p(x,y)$ where $x$ are inputs and $y$ are labels
  - Make prediction using Bayes rule to compute $p(y|x)$

- Discriminative Classifier
  - Try to predict output directly
  - Model $p(y|x)$ directly
Regression to Classification

- Thresholding on regression function does not always work
- Gaussian assumption on noise
- When the output is binary class, we may want to dry a different technique of modeling than regression
- Many modeling techniques that will better produce class category values we want for Y
- Linear Classifiers is one such method
Generative Classifier

- We can model class conditional densities using Gaussian distributions.
- If we know class conditional densities:
  - \( p(x|y=C1) \)
  - \( p(x|y=C2) \)
- We can find a decision to classify the unseen example.
Bayes Rule

\[ P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \]

\( C_1 = \text{Buys} \)
\( C_2 = \text{Doesn't Buy} \)
Generative Classifier

- Given a new data point find out posterior probability from each class and take a log ratio.
- If higher posterior probability for C1, it means new x better explained by the Gaussian distribution of C1.

\[
p(y|x) = \frac{p(x|y)p(y)}{p(x)}
\]

\[
p(y = 1|x) \propto p(x|\mu_1, \Sigma_1)p(y = 1)
\]
Naive Bayes Classifier

- Naïve Bayes Classifier a type of Generative classifier
- Compute class-conditional distribution but with conditional independence assumption
- Shown to be very useful for many classification tasks
Naive Bayes Classifier

- Conditional Independence Assumption

\[ P(X_1, X_2, \ldots, X_N | Y) = \prod_{i=1}^{N} P(X_i | Y) \]
**Naive Bayes Classifier**

\[ P(Y_k, X_1, X_2, \ldots, X_N) = P(Y_k) \prod_i P(X_i | Y_k) \]

- **Prior Probability of the Class**
- **Conditional Probability of feature given the Class**
$P(Y = y_k|X_1, X_2, ..., X_N) = \frac{P(Y = y_k)P(X_1, X_2, ..., X_N|Y = y_k)}{\sum_j P(Y = y_j)P(X_1, X_2, ..., X_N|Y = y_j)}$

$= \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_j)}$

$Y \leftarrow \operatorname{arg\,max}_{y_k} P(Y = y_k)\prod_i P(X_i|Y = y_k)$
Assignment 1

- Form a team
- Name the team
- Write a short summary about your business concept (5-10 sentences)
  - Problem addressed, your customers, your proposed idea/solution, value proposition, team
- Write 5 bullets on possible data sets you can collect to test your value proposition
- Due coming Friday @ 6pm
Assignment II

- Fill up Lean Canvas
  - [https://drive.google.com/previewtemplate?id=16uOd158UzJMX9oqwGgLo4bppzGNPmZ4fWMSV6_xBz3Z8&mode=public](https://drive.google.com/previewtemplate?id=16uOd158UzJMX9oqwGgLo4bppzGNPmZ4fWMSV6_xBz3Z8&mode=public)

- Field Assignment
  - Prepare a set of 8+ (minimum of 8) questions that you want your potential customers to answer that can help in validating your assumptions/hypotheses that you have created while filling up Lean Canvas. Questions can have qualitative or quantitative answers.
    - Example questions with quantifiable answers
      - e.g. Will you buy our product? Yes No
      - e.g. Do you think it will improve your customer service? 1 2 3 4 5
      - e.g. How important is the problem that our product solves for you? 1 2 3 4 5 6 7 8 9 10
    - Example questions with qualitative answers
      - Please tell us more about your problem?
      - What do you like about our product?

- Interview minimum of 5 potential customers
  - minimum of 3 potential customers have to be strangers (not on campus)
  - Ask the questions you have prepared
  - Interview the potential customers with any other questions that you think are relevant
  - Collect data based on questions
  - Write a few sentences about your experience with each customer

- After you are done with all interviews, write one line elevator pitch on your value proposition by filling this Mad Lib Template of Adeo Ressi
  - Our company, __insert name of company__, is developing __a defined offering__ to help __a defined audience__ __solve a problem__ with __secret sauce__.

- Due Feb 18th Monday @ 6pm