



MAN VS. MACHINE

How IBM Built a *Jeopardy!* Champion

15.071x – The Analytics Edge

A Grand Challenge



- In 2004, IBM Vice President Charles Lickel and co-workers were having dinner at a restaurant
- All of a sudden, the restaurant fell silent
- Everyone was watching the game show *Jeopardy!* on the television in the bar
- A contestant, Ken Jennings, was setting the record for the longest winning streak of all time (75 days)

A Grand Challenge



- Why was everyone so interested?
 - *Jeopardy!* is a quiz show that asks complex and clever questions (puns, obscure facts, uncommon words)
 - Originally aired in 1964
 - A huge variety of topics
 - Generally viewed as an impressive feat to do well
- No computer system had ever been developed that could even come close to competing with humans on *Jeopardy!*

A Tradition of Challenges



- IBM Research strives to push the limits of science
 - Have a tradition of inspiring and difficult challenges
- Deep Blue – a computer to compete against the best human chess players
 - A task that people thought was restricted to human intelligence
- Blue Gene – a computer to map the human genome
 - A challenge for computer speed and performance

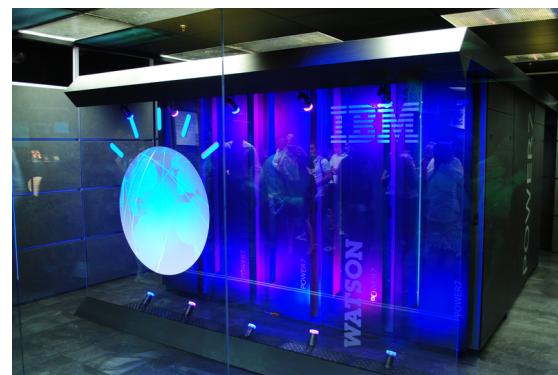
The Challenge Begins



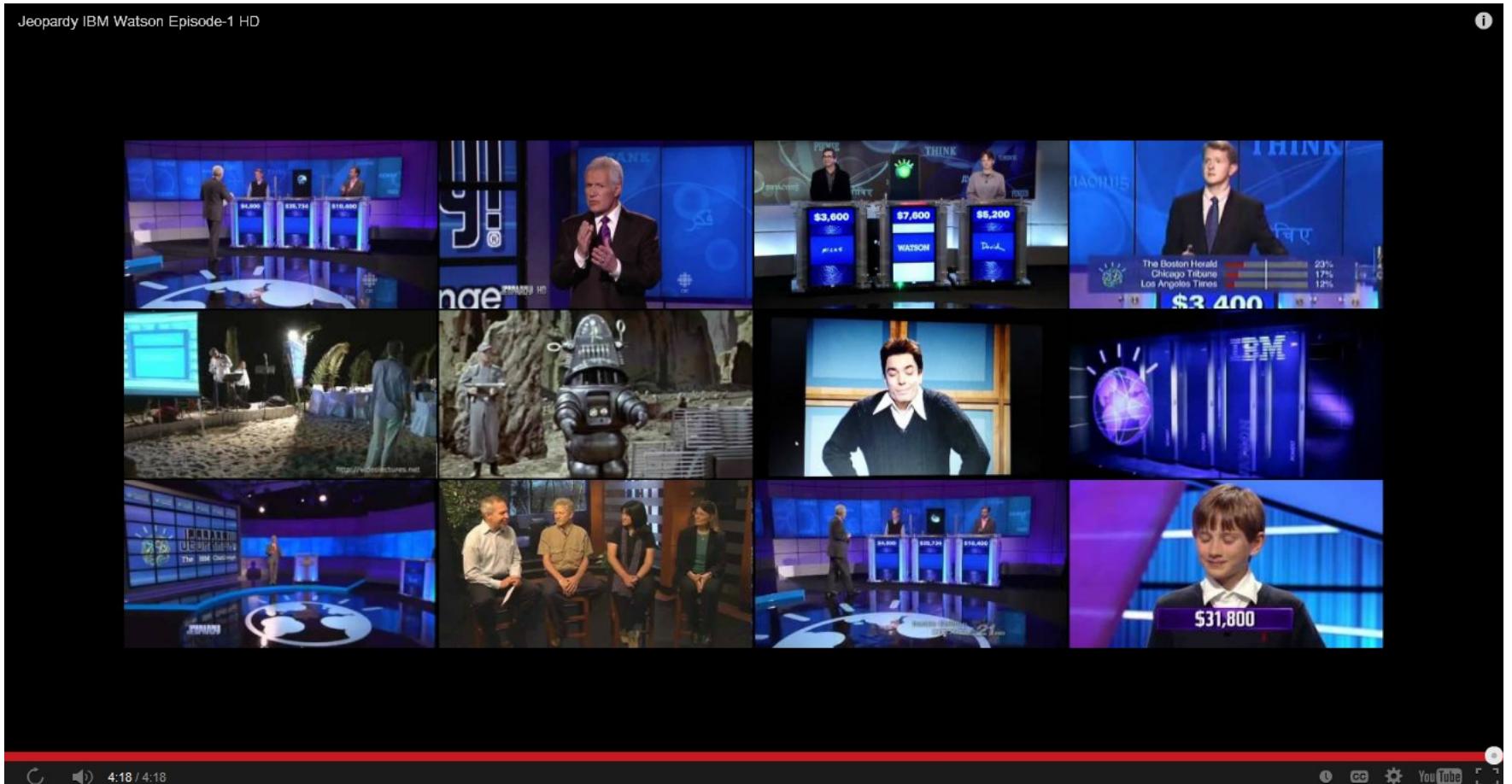
- In 2005, a team at IBM Research started creating a computer that could compete at *Jeopardy!*
 - No one knew how to beat humans, or if it was even possible
- Six years later, a two-game exhibition match aired on television
 - The winner would receive \$1,000,000

The Contestants

- Ken Jennings
 - Longest winning streak of 75 days
- Brad Rutter
 - Biggest money winner of over \$3.5 million
- Watson
 - A supercomputer with 3,000 processors and a database of 200 million pages of information



The Match Begins



The Game of *Jeopardy!*

- Three rounds per game
 - Jeopardy
 - Double Jeopardy (dollar values doubled)
 - Final Jeopardy (wager on response to one question)
- Each round has five questions in six categories
 - Wide variety of topics (over 2,500 different categories)
- Each question has a dollar value – the first to buzz in and answer correctly wins the money
 - If they answer incorrectly they lose the money



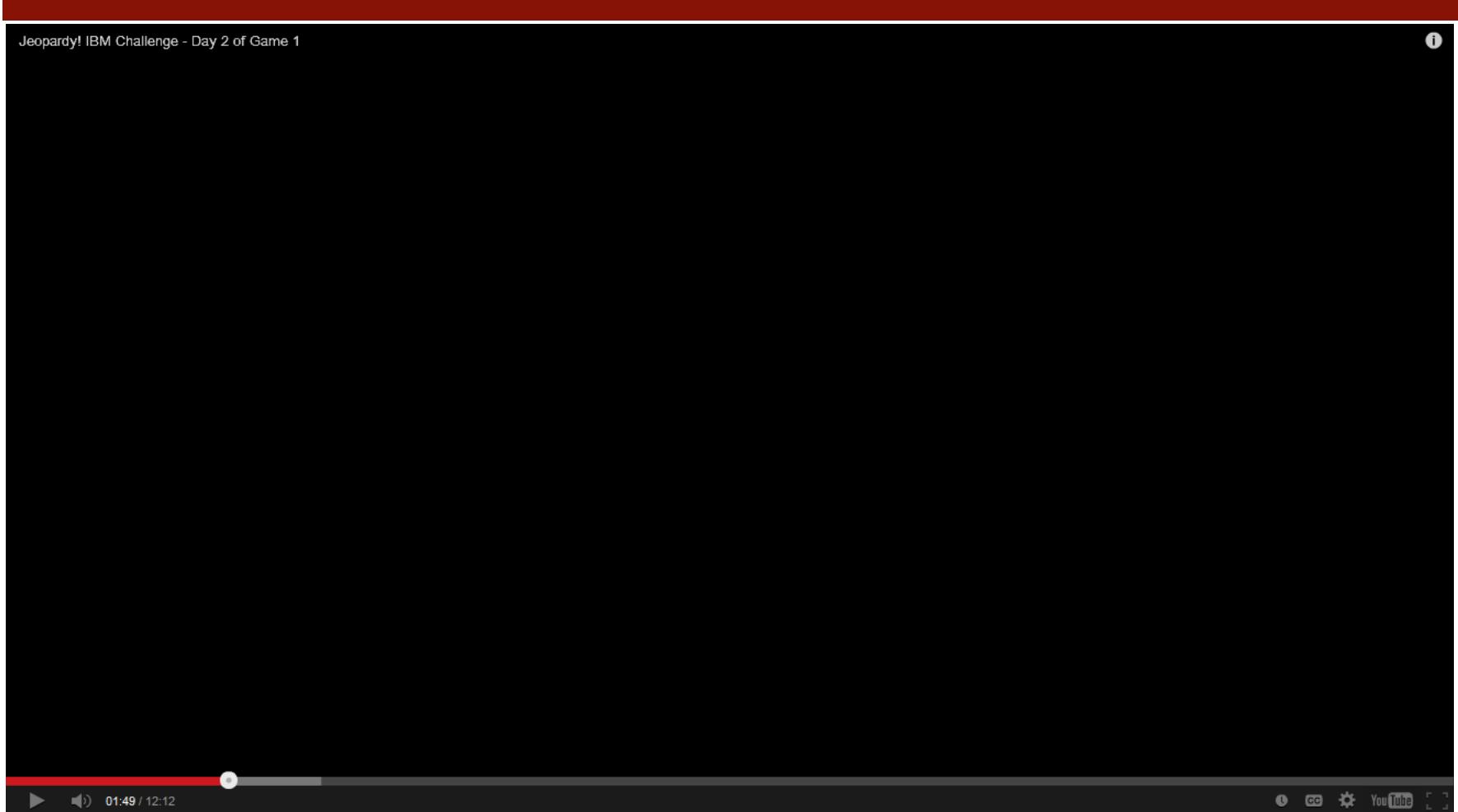
Example Round

THE DINOSAURS	NOTABLE WOMEN	OXFORD ENGLISH DICTIONARY	NAME THAT INSTRUMENT	BELGIUM	COMPOSERS BY COUNTRY
\$200	\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000	\$1000

Jeopardy! Questions

- Cryptic definitions of categories and clues
- Answer in the form of a question
 - Q: Mozart's last and perhaps most powerful symphony shares its name with this planet.
 - A: What is Jupiter?
 - Q: Smaller than only Greenland, it's the world's second-largest island.
 - A: What is New Guinea?

Watson Playing Jeopardy



Why is Jeopardy Hard?



- Wide variety of categories, purposely made cryptic
- Computers can easily answer precise questions
 - What is the square root of $(35672-183)/33$?
- Understanding natural language is hard
 - Where was Albert Einstein born?
 - Suppose you have the following information:
“One day, from his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of his birthplace.”
 - Ulm? Otto?

A Straightforward Approach



- Let's just store answers to all possible questions
- This would be impossible
 - An analysis of 200,000 previous questions yielded over 2,500 different categories
- Let's just search Google
 - No links to the outside world permitted
 - It can take considerable skill to find the right webpage with the right information

Using Analytics



- Watson received each question in text form
 - Normally, players see and hear the questions
- IBM used analytics to make Watson a competitive player
- Used over 100 different techniques for analyzing natural language, finding hypotheses, and ranking hypotheses

Watson's Database and Tools

- A massive number of data sources
 - Encyclopedias, texts, manuals, magazines, Wikipedia, etc.
- Lexicon
 - Describes the relationship between different words
 - Ex: “Water” is a “clear liquid” but not all “clear liquids” are “water”
- Part of speech tagger and parser
 - Identifies functions of words in text
 - Ex: “Race” can be a verb or a noun
 - He won the race by 10 seconds.
 - Please indicate your race.

How Watson Works

- Step 1: Question Analysis
 - Figure out what the question is looking for
- Step 2: Hypothesis Generation
 - Search information sources for possible answers
- Step 3: Scoring Hypotheses
 - Compute confidence levels for each answer
- Step 4: Final Ranking
 - Look for a highly supported answer

Step 1: Question Analysis

- What is the question looking for?
- Trying to find the Lexical Answer Type (LAT) of the question
 - Word or noun in the question that specifies the type of answer
- Ex: “Mozart’s last and perhaps most powerful symphony shares its name with **this planet**.”
- Ex: “Smaller than only Greenland, **it’s** the world’s second-largest island.”

Step 1: Question Analysis

- If we know the LAT, we know what to look for
- In an analysis of 20,000 questions
 - 2,500 distinct LATs were found
 - 12% of the questions do not have an explicit LAT
 - The most frequent 200 explicit LATs cover less than 50% of the questions
- Also performs **relation detection** to find relationships among words, and **decomposition** to split the question into different clues

Step 2: Hypothesis Generation



- Uses the question analysis from Step 1 to produce candidate answers by searching the databases
- Several hundred candidate answers are generated
- Ex: “Mozart’s last and perhaps most powerful symphony shares its name with **this planet**.”
 - Candidate answers: Mercury, Earth, Jupiter, etc.

Step 2: Hypothesis Generation

- Then each candidate answer plugged back into the question in place of the LAT is considered a hypothesis
 - Hypothesis 1: “Mozart’s last and perhaps most powerful symphony shares its name with **Mercury**.”
 - Hypothesis 2: “Mozart’s last and perhaps most powerful symphony shares its name with **Jupiter**.”
 - Hypothesis 3: “Mozart’s last and perhaps most powerful symphony shares its name with **Earth**.”

Step 2: Hypothesis Generation



- If the correct answer is not generated at this stage, Watson has no hope of getting the question right
- This step errors on the side of generating a lot of hypotheses, and leaves it up to the next step to find the correct answer

Step 3: Scoring Hypotheses

- Compute *confidence levels* for each possible answer
 - Need to accurately estimate the probability of a proposed answer being correct
 - Watson will only buzz in if a confidence level is above a threshold
- Combines a large number of different methods

Lightweight Scoring Algorithms

- Starts with “lightweight scoring algorithms” to prune down large set of hypotheses
- Ex: What is the likelihood that a candidate answer is an instance of the LAT?
 - If this likelihood is not very high, throw away the hypothesis
- Candidate answers that pass this step proceed the next stage
 - Watson lets about 100 candidates pass into the next stage

Scoring Analytics

- Need to gather supporting evidence for each candidate answer
- Passage Search
 - Retrieve passages that contain the hypothesis text
 - Let's see what happens when we search for our hypotheses on Google
 - Hypothesis 1: "Mozart's last and perhaps most powerful symphony shares its name with **Mercury**."
 - Hypothesis 2: "Mozart's last and perhaps most powerful symphony shares its name with **Jupiter**."

Passage Search



About 938,000 results (0.55 seconds)

[Mercury: Mozart's Jupiter Symphony - The Front Row](#)

www.thefrontrow.org/.../1349112026-Mercury-Mozarts-Jupiter-Sympho... ▾

Oct 1, 2012 - Antoine Plante, artistic director of the period-instruments group **Mercury** - The Orchestra Redefined, talks about the program of **symphonies** and ...

[Mozarts Jupiter Symphony | Mercury \(formerly Mercury Baro...](#)

www.artshound.com › MUSIC ▾

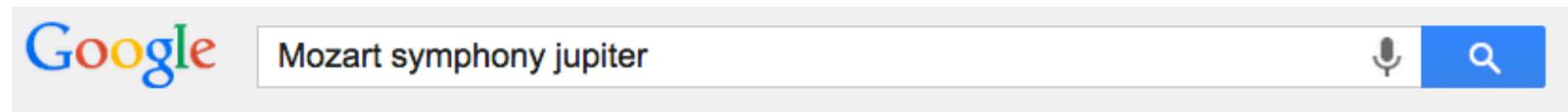
Opening the **Mercury** season at the Wortham Center's Cullen Theatre on Saturday, October 6, 2012 will be a program featuring **Mozart's "Jupiter" Symphony**.

[Event - Mozart's "Jupiter" Symphony Mercury Houston - The ...](#)

mercuryhouston.org/events/7/ ▾

Mercury combines the forces of Haydn and **Mozart** for a memorable concert event, highlighted by **Mozarts** iconic Jupiter **Symphony**. A wonderful way to kick off ...

Passage Search



Web Images Maps Shopping Videos More ▾ Search tools

About 1,440,000 results (0.31 seconds)

[Symphony No. 41 \(Mozart\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Symphony_No._41_\(Mozart\)](http://en.wikipedia.org/wiki/Symphony_No._41_(Mozart)) ▾

It was the last **symphony** that he composed, and also the longest. The work is nicknamed the **Jupiter Symphony**. This name stems not from **Mozart** but rather was ...

[Instrumentation](#) - [Composition and premiere](#) - [Movements](#) - [Notes](#)

[W. A. Mozart - Symphony No. 41 "Jupiter" in C major \(Harmon...](#)

www.youtube.com/watch?v=zK5295yEQMQ ▾

Feb 11, 2012 - W. A. Mozart - Symphony No. 41 "Jupiter" in C major, K. 551 (1788): 1. Allegro vivace, 4/4 2. Andante cantabile, 3/4 in F major 3. Menuetto: ...

[Mozart - Symphony No. 41 in C, K. 551 \[complete\] \(Jupiter\) - ...](#)

www.youtube.com/watch?v=bnK3kh8ZEgA ▾

Feb 21, 2012 - Wolfgang Amadeus Mozart completed his **Symphony** No. 41 in C major, K. 551, on 10 August 1788. It was the last **symphony** that he composed ...

Scoring Analytics

- Determine the degree of certainty that the evidence supports the candidate answers
- More than 50 different scoring components
- Ex: Temporal relationships
 - “In 1594, he took a job as a tax collector in Andalusia”
 - Two candidate answers: Thoreau and Cervantes
 - Thoreau was not born until 1817, so we are more confident about Cervantes

Step 4: Final Merging and Ranking



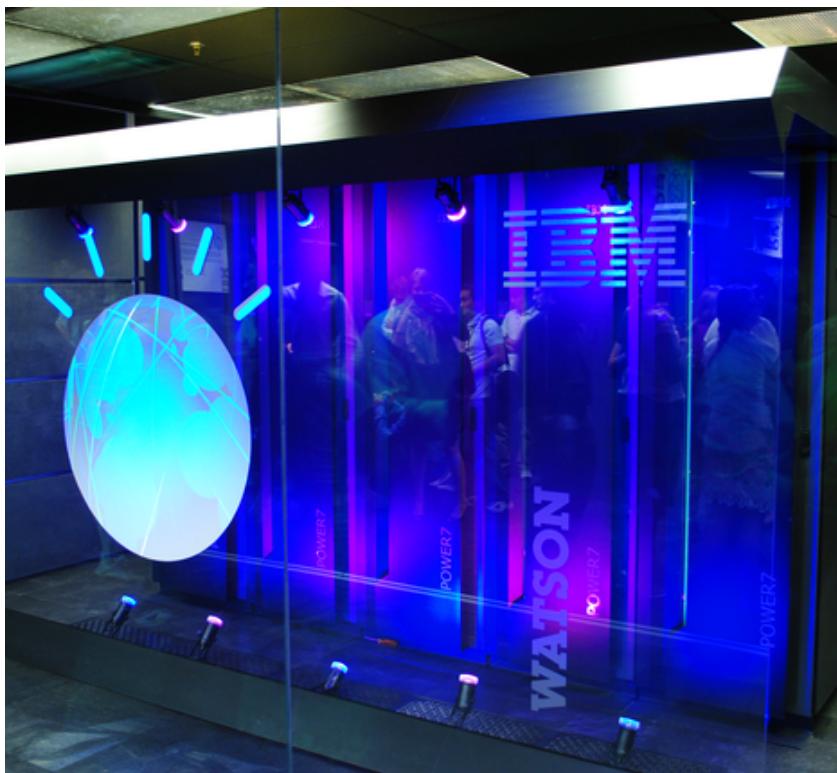
- Selecting the single best supported hypothesis
- First need to merge similar answers
 - Multiple candidate answers may be equivalent
 - Ex: “Abraham Lincoln” and “Honest Abe”
 - Combine scores
- Rank the hypotheses and estimate confidence
 - Use predictive analytics

Ranking and Confidence Estimation



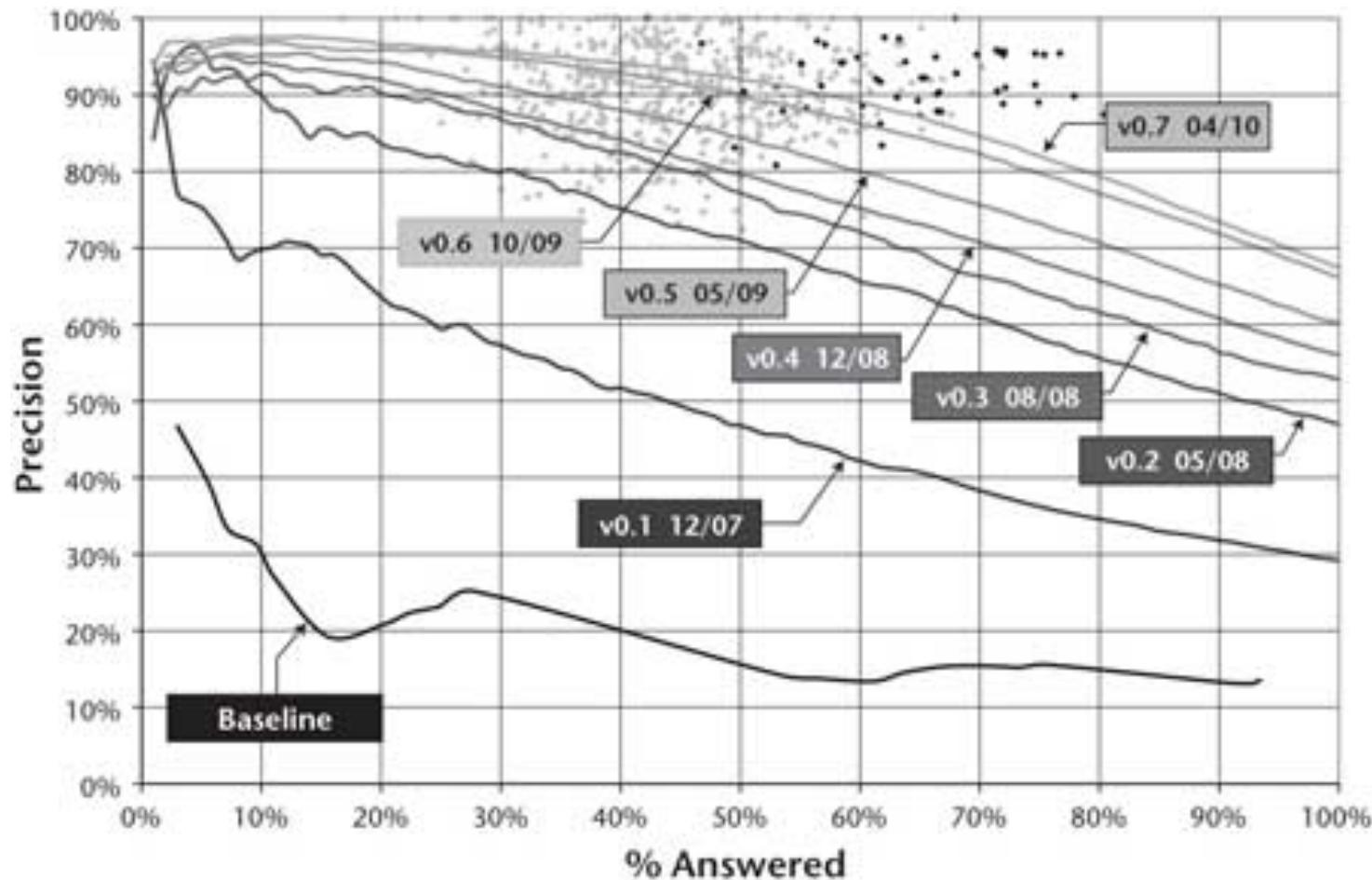
- Training data is a set of historical *Jeopardy!* questions
- Each of the scoring algorithms is an independent variable
- Use logistic regression to predict whether or not a candidate answer is correct, using the scores
- If the confidence for the best answer is high enough, Watson buzzes in to answer the question

The Watson System



- Eight refrigerator-sized cabinets
- High speed local storage for all information
- Originally took over two hours to answer one question
 - This had to be reduced to 2-6 seconds

Progress from 2006 - 2010



Let the games begin!



- The games were scheduled for February 2011
- Two games were played, and the winner would be the contestant with the highest winnings over the two games

The Jeopardy Challenge



The Results

	Ken Jennings	Brad Rutter	Watson
Game 1	\$4,800	\$10,400	\$35,734
Game 2	\$19,200	\$11,200	\$41,413
Total	\$24,000	\$21,600	\$77,147

What's Next for Watson

- Apply to other domains
 - Watson is ideally suited to answering questions which cover a wide range of material and often have to deal with inconsistent or incomplete information
- Medicine
 - The amount of medical information available is doubling every 5 years and a lot of the data is unstructured
 - Cancer diagnosis and selecting the best course of treatment
 - MD Anderson and Memorial Sloan-Kettering Cancer Centers

The Analytics Edge



- Combine many algorithms to increase accuracy and confidence
 - Any one algorithm wouldn't have worked
- Approach the problem in a different way than how a human does
 - Hypothesis generation
- Deal with massive amounts of data, often in unstructured form
 - 90% of data is unstructured