Computing the Probabilities of Closing of 10b-5 Securities Class Action Cases

Steve Hillmer and Prakash P. Shenoy

KU CENTER FOR BUSINESS ANALYTICS RESEARCH
The University of Kansas
School of Business

CBAR Seminar
January 31, 2014

Outline

- Problem
- Data
- Model and Results
- Limitations
- Future Work

Problem

- Given a new 10b-5 securities class-action case, filed in a Federal court, where lead plaintiff and lead plaintiff counsel have been appointed, and a consolidated amended complaint has been filed, what is the probability that the case will be dismissed?
- This problem is relevant to D&O insurance companies from a claims perspective. During 1997–2013, securities class action cases have settled for a total of $73 billion, not including $15 billion in plaintiff lawyers fees and an equivalent amount in defendant lawyers fees.

Problem

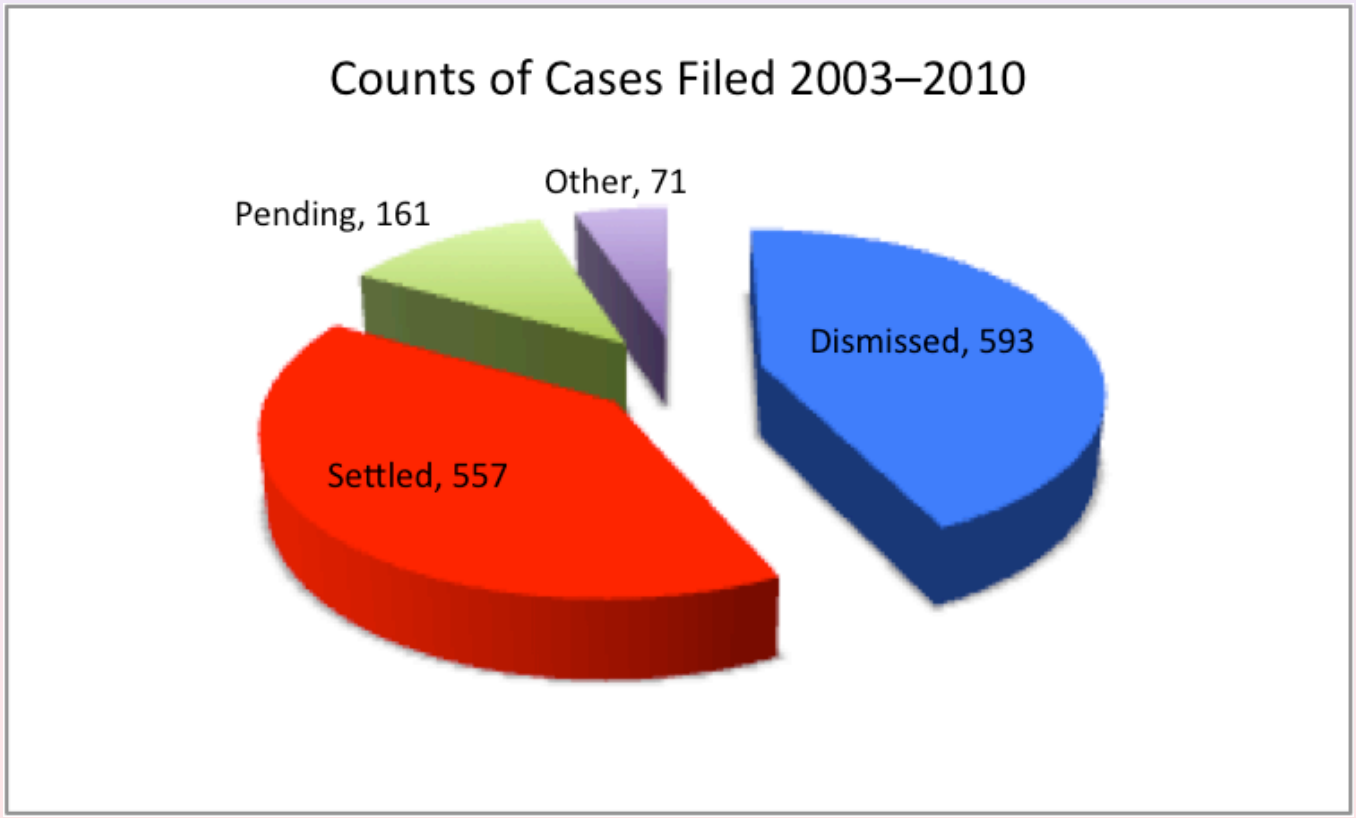
- This is part of a related problem from an underwriting perspective: Given a potential D&O customer,
  - What is the probability that a class-action complaint will be filed?
  - Assuming that a class-action complaint is filed, what is the probability that it will be dismissed?
  - Assuming that it is not dismissed, what are the potential settlement amounts?

Outline

- Problem
- Data
- Model and Results
- Limitations
- Future Work

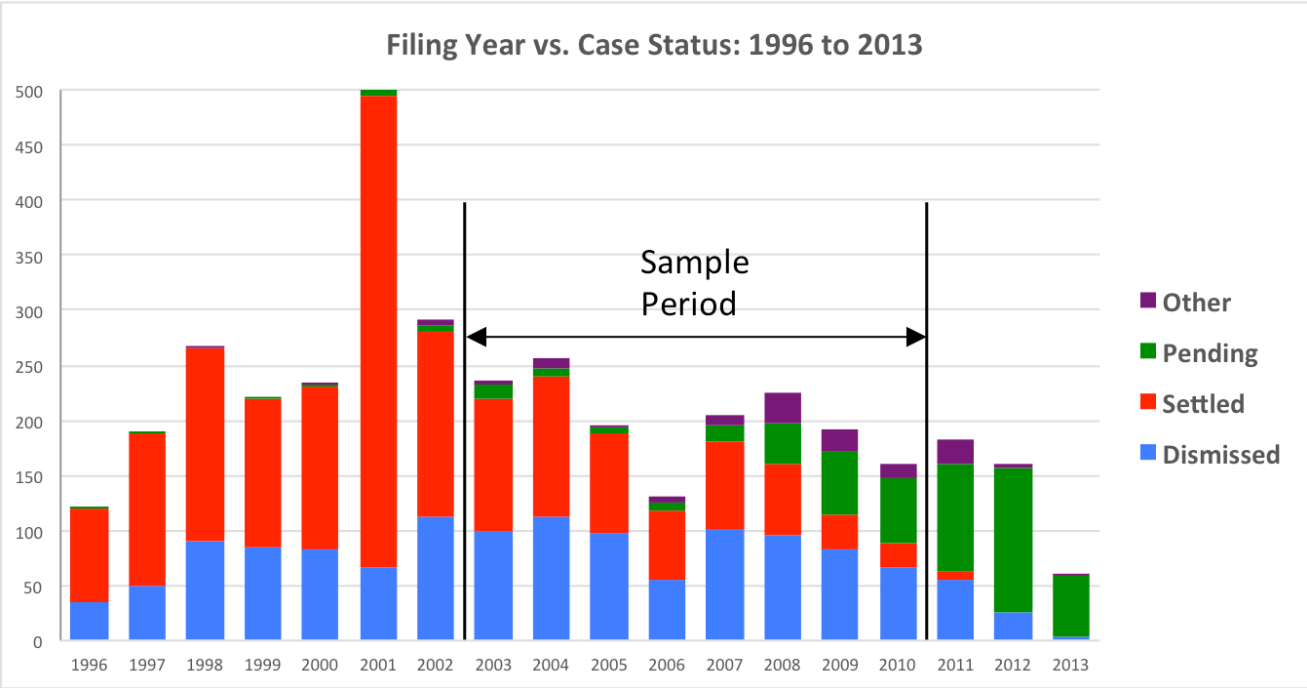
Data

- Closed 10b-5 securities class-action cases filed during 2003–2010. 1,150 such cases.



Data

- Why 2003–2010? Sarbanes-Oxley Act of 2002 has changed the characteristics of 10b-5 cases. Many of the cases in the period 2011 onwards are still pending.



Data

Sources of data:

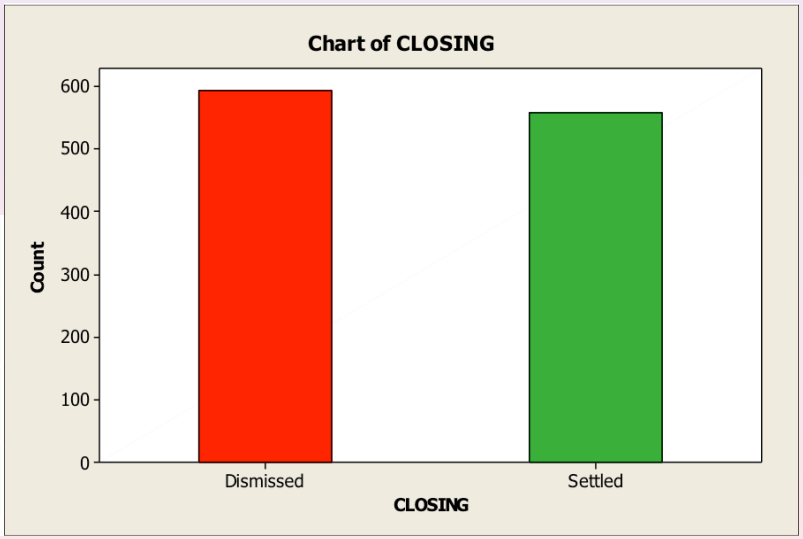
1. Advisen's Master Significant Cases and Actions database (MSCAd) for case data
2. Stanford's Class Action Clearinghouse to verify case details
3. COMPUSTAT for financial information about securities

Data

The class variable is

- CLOSING: with two possible values: dismissed or settled

CLOSING	Count	Percent
Dismissed	593	52%
Settled	557	48%
All	1150	



Data

Predictor variables (in our model):

Name	*Description*	*Possible Values*	*Type*	*Missing?*
GAAP	whether violations of generally accepted accounting principles is alleged, or not	1, 0	Nominal	
SEC_11	whether case involves filing false documents with the SEC, or not	1, 0	Nominal	
INST_INV	whether lead plaintiff is an institution, or individuals	1, 0	Nominal	
RESTATED_FIN	if re-stated financials are involved, or not	1, 0	Nominal	
BANKRUPTCY	whether bankruptcy is involved in the complaint, or not	1, 0	Nominal	
ONE_DAY_DROP	largest one-day drop in the price of the security, adjusted for market, during one year preceding the filing of the first complaint		Numeric	56%

Data

Predictor variables we considered (but not in model):

Name	*Description*	*Possible Values*	*Type*	*Missing?*
LPLF_Type	Lead Plaintiff Law Firm Type	U, M, L, N, X	Nominal	11%
INSIDER_TRADING	whther insider trading is alleged, or not	1, 0	Nominal	
TRANSACTIONAL	whether the case involves sale/merger/ acquisition, or not	1, 0	Nominal	
LADDERING	whether laddering is alleged, or not	1, 0	Nominal	
3RD_PARTY_BANKRUPTCY	whether the complaint is related to the bankruptcy of a 3rd party, or not	1, 0	Nominal	
IPO	whether initial public offering is involved, or not	1, 0	Nominal	
PO	whether secondary public offering is involved, or not	1, 0	Nominal	

Data

Predictor variables we considered (but not in model):

Name	*Description*	*Possible Values*	*Type*	*Missing?*
ERISA	whether case involves a violation of the Employee Retirement Income Securities Act, or not	1, 0	Nominal	
FCPA	whether case involves Foreign Corrupt Practices Act, or not	1, 0	Nominal	
DOJ_INV	if case involves Dept. of Justice investigation, or not	1, 0	Nominal	
SEC_INV	whether the case involves investigation by the SEC Commission, or not	1, 0	Nominal	
TOTAL_ASSET	total assets of the company around the filing date, quarterly		Numeric	26%
TOTAL_REV	total revenue of the company around the filing date, quarterly		Numeric	26%

Outline

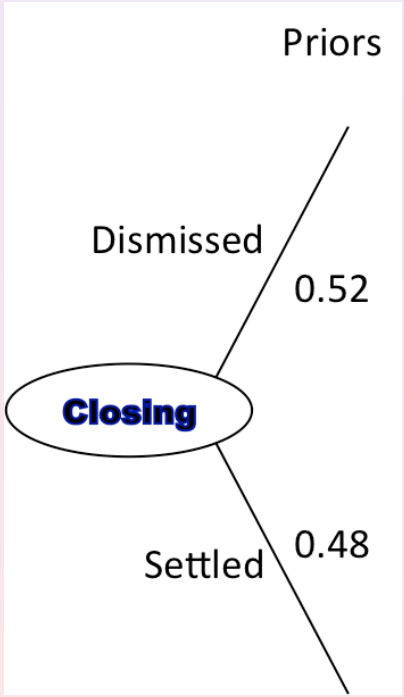
- Problem
- Data
- Model and Results
  - What are naïve Bayes (NB) models?
  - How does one use a NB model?
  - Why NB?
  - What is our NB model?
    - What are the predictors?
    - What are the parameters?
  - How good is our NB model?
  - What are the relative influences of the predictor variables?
- Limitations
- Future Work

Model and Results

- We use a naïve Bayes model to compute posterior probabilities of *CLOSING* given observed values of a subset of variables
- What are naïve Bayes models?

Model and Results: What are NB Models?

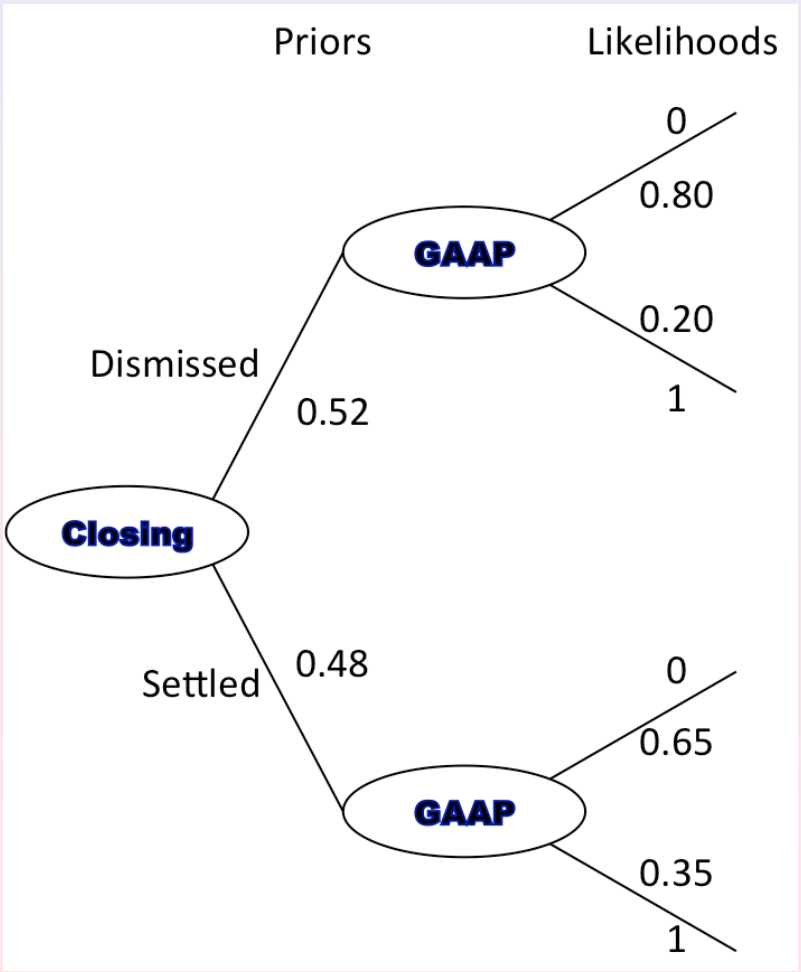
- Let's construct a naïve Bayes model with, e.g., CLOSING, GAAP, and INST_INV



CLOSING	Count	Percent
Dismissed	593	52%
Settled	557	48%
All	1150	

Model and Results: What are NB Models?

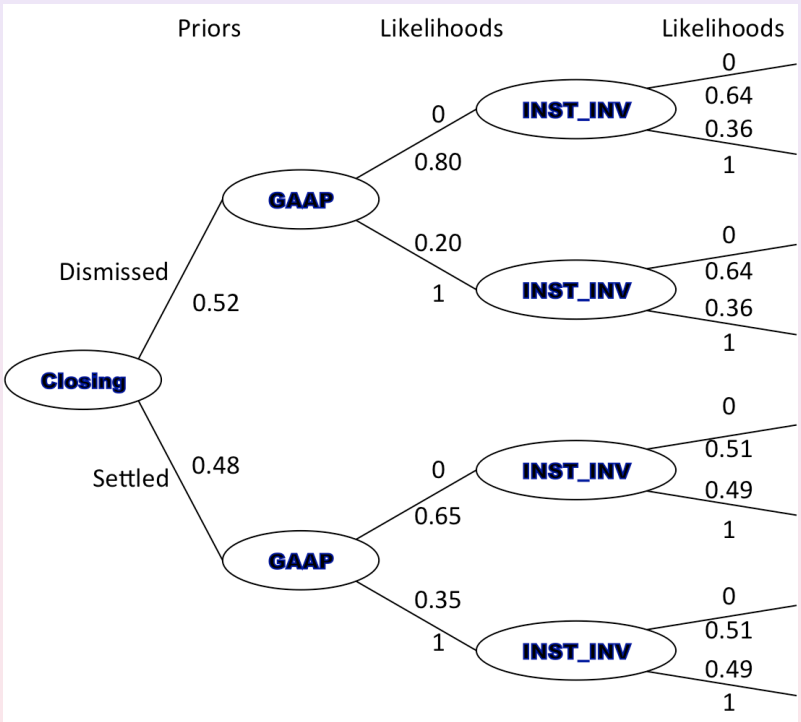
- Adding GAAP:



Rows: GAAP	Columns: CLOSING		
	Dismissed	Settled	All
0	476	364	840
	80%	65%	73%
1	117	193	310
	20%	35%	27%
All	593	557	1150

Model and Results: What are NB Models?

- Adding INST_INV:



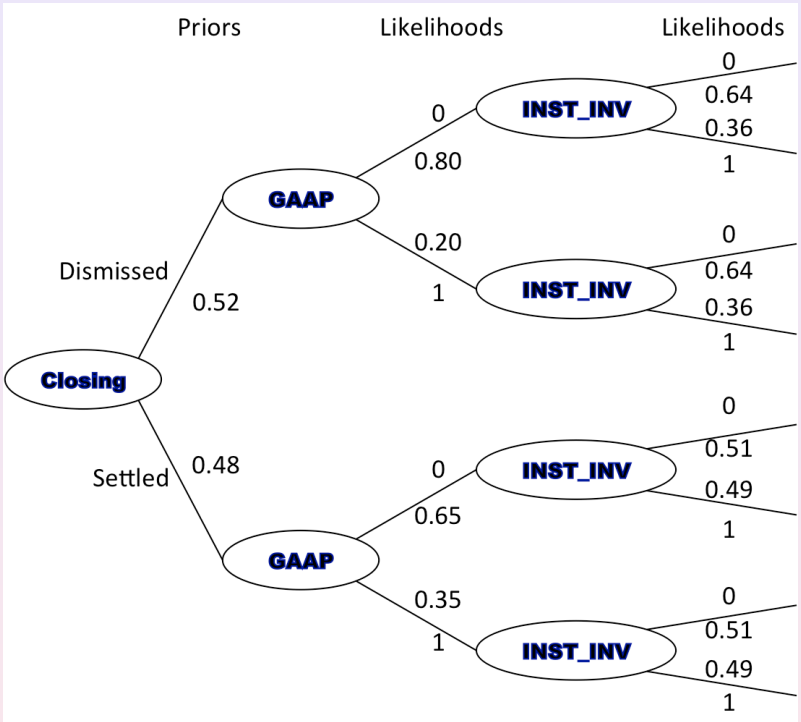
Rows: INST_INV Columns: CLOSING

	Dismissed	Settled	All
0	378	282	660
	64%	51%	57%
1	215	275	490
	36%	49%	43%
All	593	557	1150

- We are assuming that given *CLOSING*, probabilities of *INST_INV* are independent of *GAAP*

Model and Results: What are NB Models?

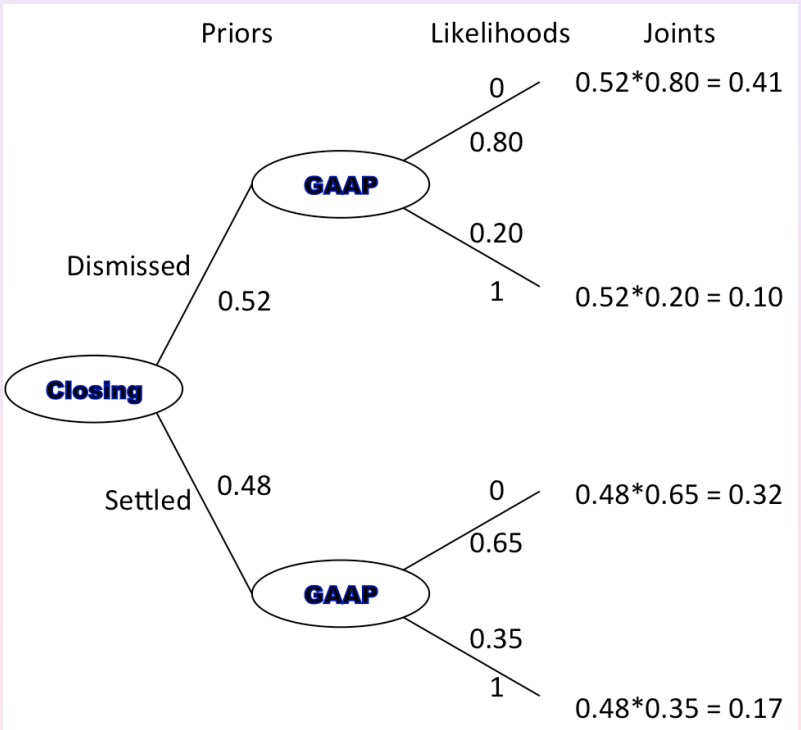
- This model has only 5 parameters (instead of 7)



- A naïve Bayes model with 10 predictor variables would have only 21 parameters, instead of $1 + 2 + 4 + 8 + \ldots + 1{,}024 = 2{,}047$ without the independence assumption.

Model and Results: How does one use a NB Model?

- Consider the model with CLOSING and GAAP:

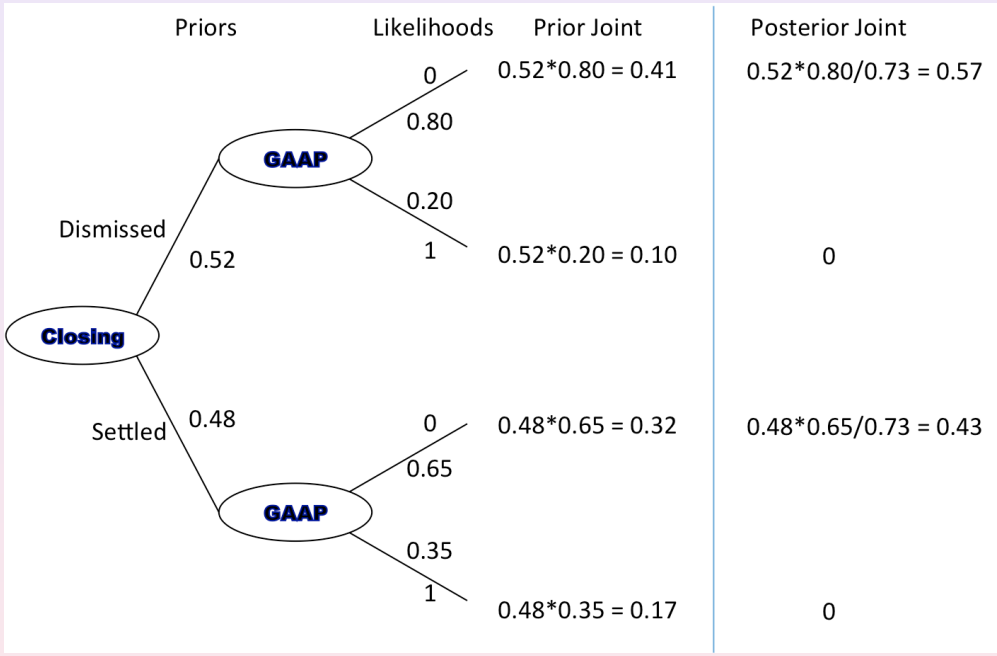


Rows: GAAP	Columns: CLOSING		
	Dismissed	Settled	All
0	476	364	840
	41%	32%	73%
1	117	193	310
	10%	17%	27%
All	593	557	1150
	52%	48%	100%

- Multiplication of *priors* and *likelihoods* gives us joints that add to 1.

Model and Results: How does one use a NB Model?

- Suppose $GAAP = 0$. What is the posterior $Pr(dismissed)$?



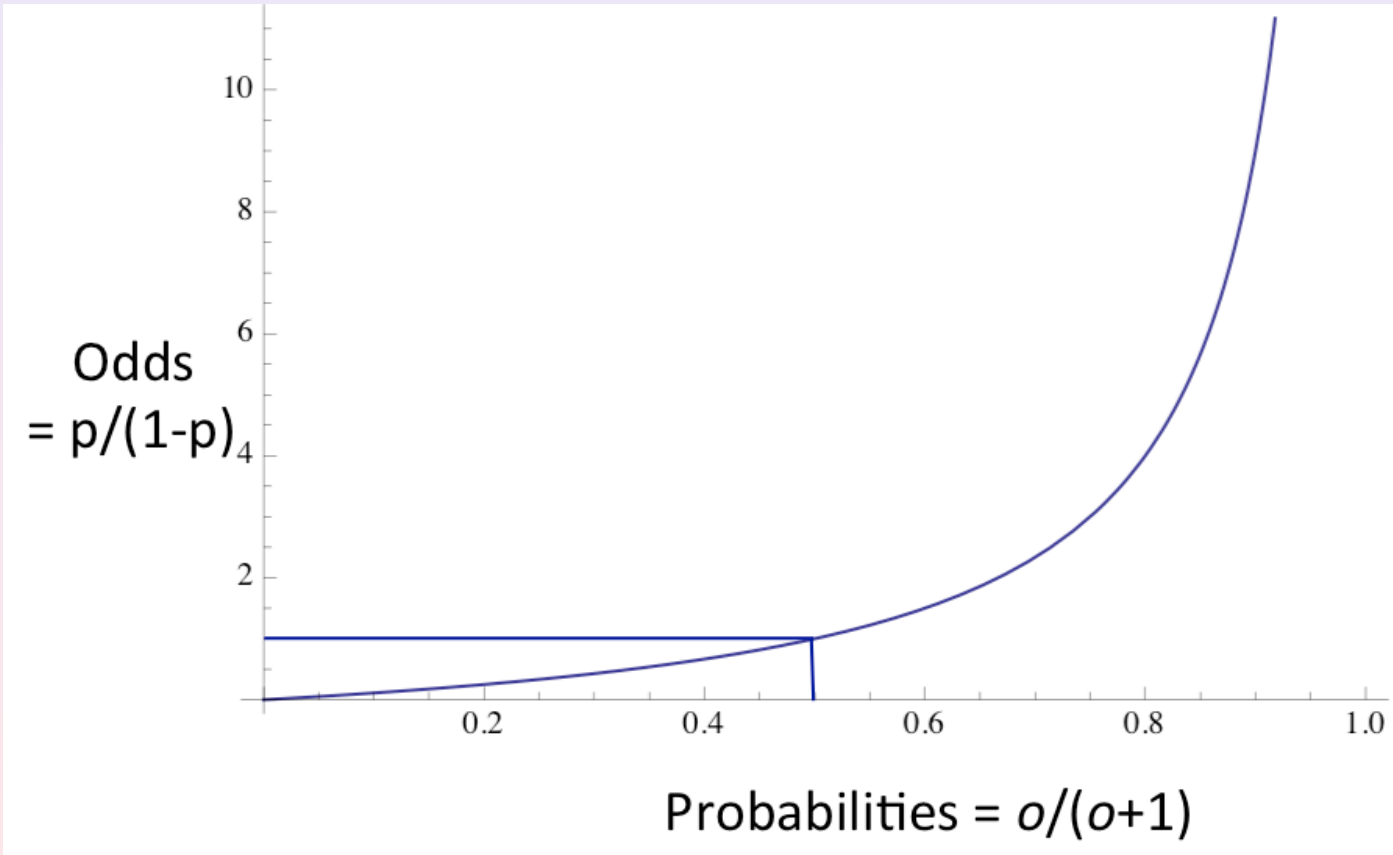
- This way of computing posterior probabilities is not tractable when we have many predictors
- It is easier to use odds and likelihood ratios

Model and Results: Probabilities and Odds

What are odds?

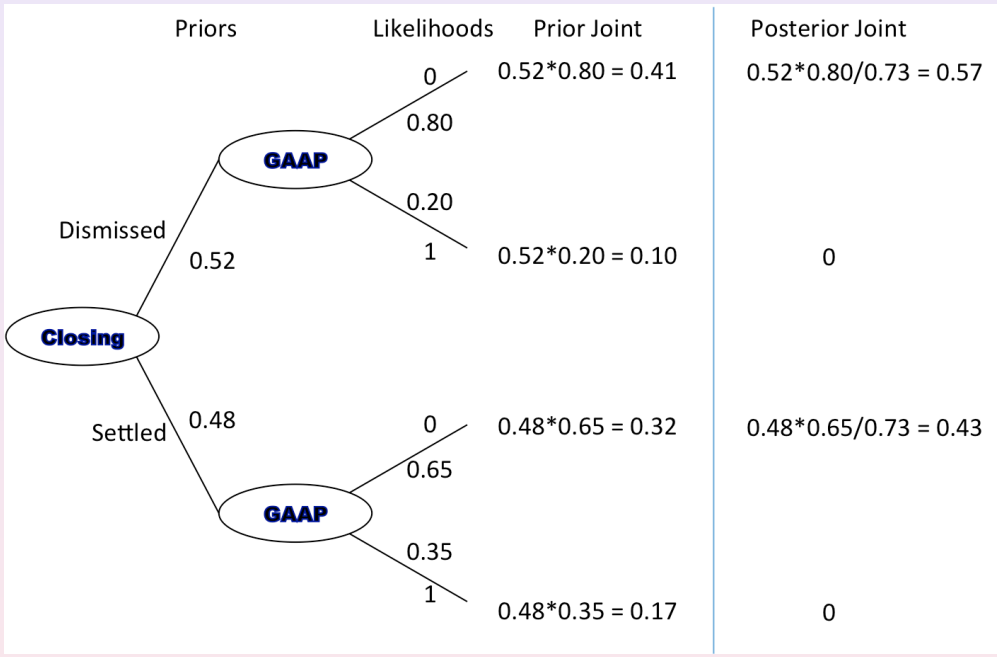
- $Pr(dismissed) = 0.52$ is equivalent to $O(dismissed) = \frac{0.52}{1-0.52} = 1.08$
- Probabilities are on a scale from 0 to 1. Odds are on a scale from 0 to ∞
- $O(dismissed) = 1.08$ is equivalent to $Pr(dismissed) = \frac{1.08}{1.08+1} = 0.52$.
- $O(dismissed) = 1$, means $Pr(dismissed) = 0.5$
  $O(dismissed) > 1$, means $Pr(dismissed) > 0.5$
  $O(dismissed) < 1$, means $Pr(dismissed) < 0.5$

Model and Results: Probabilities and Odds



Model and Results: How does one use a NB Model?

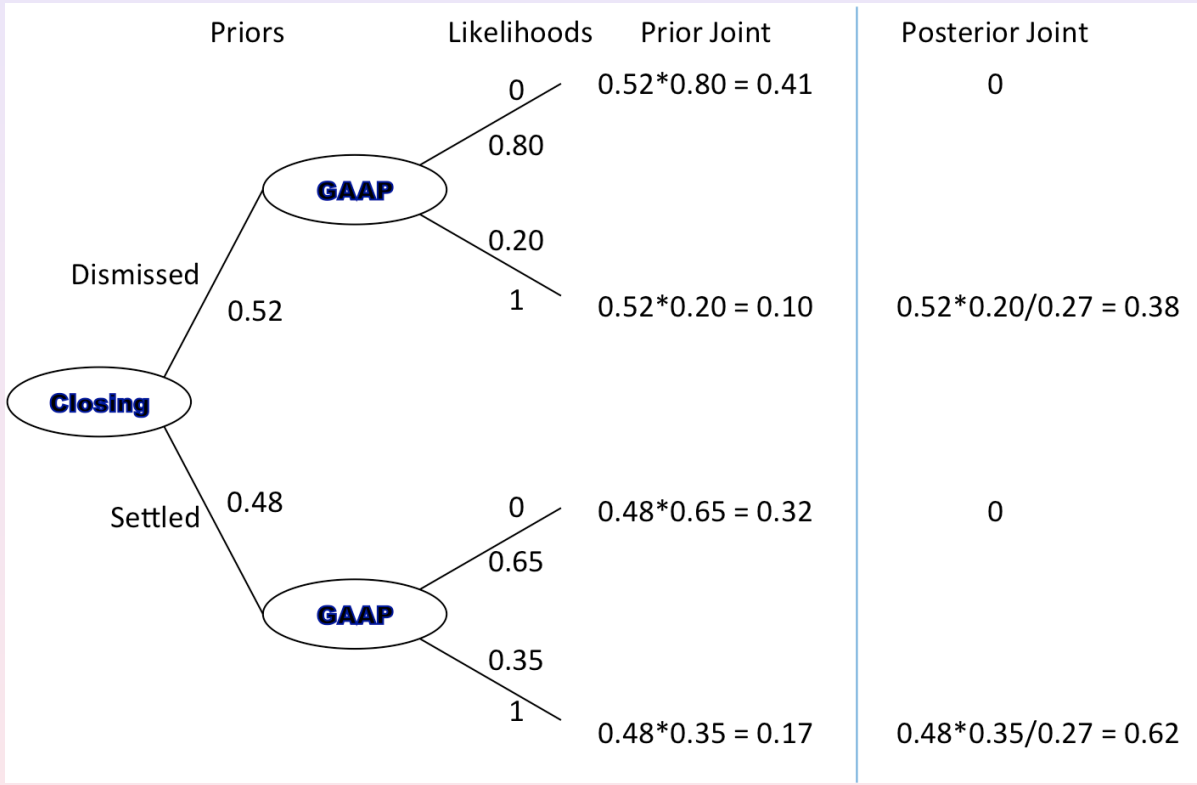
- Suppose $GAAP = 0$. What are the posterior $O(dismissed)$?



- Posterior $O(dismissed) = \frac{0.52 \cdot 0.80}{0.48 \cdot 0.65} = \frac{0.52}{0.48} \cdot \frac{0.80}{0.65}$
  $= O(dismissed) \times LR(GAAP = 0) = 1.08 \cdot 1.23 = 1.31.$
  $\therefore$ Posterior $Pr(dismissed) = 0.57.$

Model and Results: How does one use a NB Model?

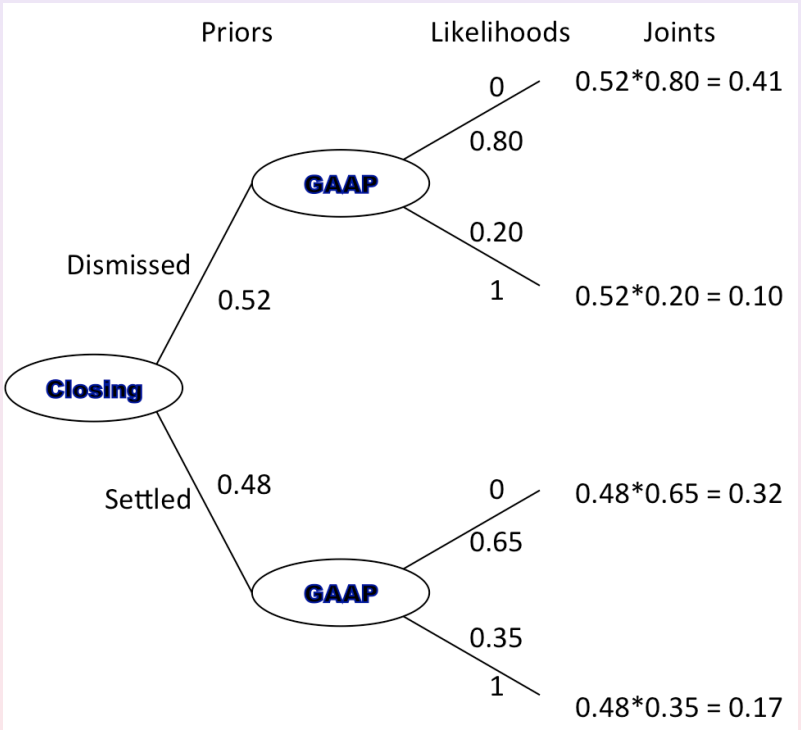
- Suppose $GAAP = 1$. What are posterior odds of *dismissed*?



- Posterior $O(dismissed) = \frac{0.52 \cdot 0.20}{0.48 \cdot 0.35} = \frac{0.52}{0.48} \cdot \frac{0.20}{0.35} =$
  $O(dismissed) \times LR(GAAP = 1) = 1.08 \cdot 0.57 = 0.62.$
  $\therefore$ Posterior $Pr(dismissed) = 0.38.$

Model and Results: How does one use a NB Model?

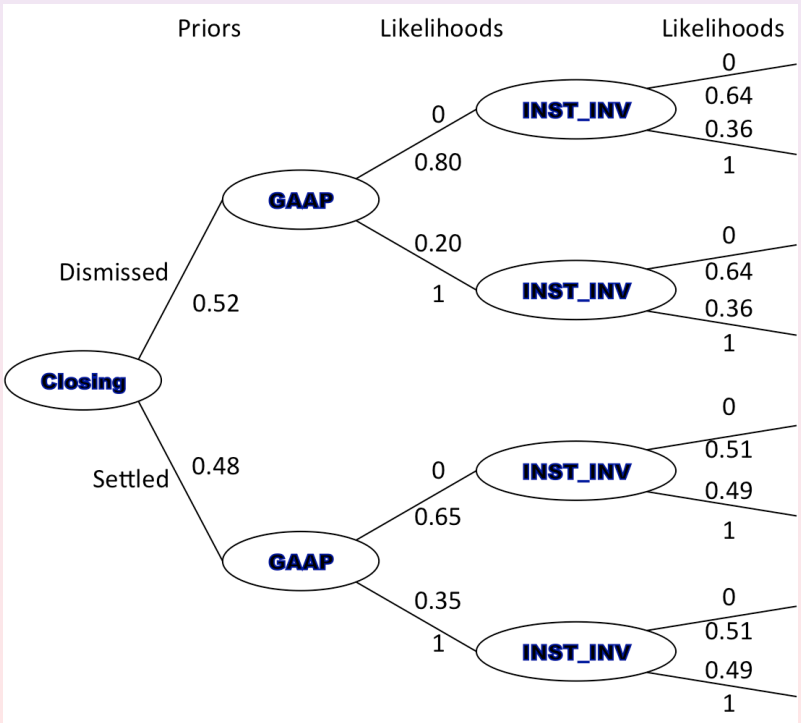
- Suppose GAAP is unknown. What are posterior odds of *dismissed*?



- Posterior $O(dismissed)$ = Prior $O(dismissed) = \frac{0.52}{0.48}$
  $LR(GAAP = unknown) = \frac{1}{1} = 1$

Model and Results: How does one use a NB Model?

- Easy to compute posterior odds (or probabilities) of *dismissed*:
  Posterior odds = Prior odds $\times$ *Likelihood ratio* of evidence
- Suppose $GAAP = 1$, and $INST\_INV = 0$.
  Posterior $O(dismissed)$ = $O(dismissed) \times LR(GAPP = 1)$
  $\times\ LR(INST\_INV = 0)$
  $= \frac{0.52}{0.48} \times \frac{0.20}{0.35} \times \frac{0.64}{0.51} = 0.78$. $\therefore Pr(dismissed) = 0.43$.



Model and Results: Why use NB?

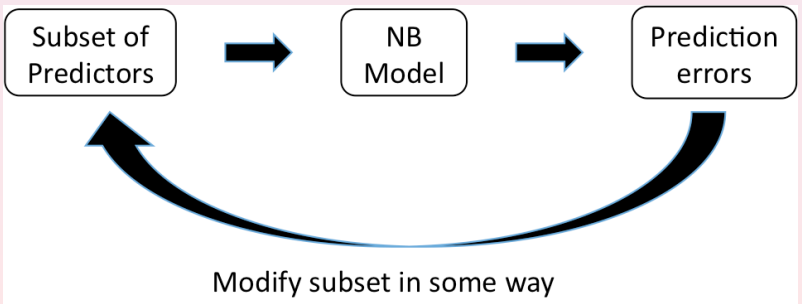
Why use naïve Bayes?

- Provides probabilities of dismissed and settled;
- Can be used even if some predictor variables have missing values;
- Variables can be numeric or nominal;
- Simple—has very few parameters (# parameters is linear in # predictor variables);
- Robust—predicts well even if the independence assumption of the model is violated;

Model and Results: What are the predictors?

Which subset of predictor variables provides the "best" naïve Bayes model?

- By "best", we mean a model that has the lowest out-of-sample prediction errors
- Given 19 predictor variables, we have $2^{19} - 1 = 524,287$ non-empty subsets—too many to enumerate
- We did a search using several methods: best first, random, etc.



Model and Results: What are the predictors?

Which subset of predictor variables results in a good naïve Bayes model?

- Answer: The subset consisting of
  1. GAAP (violations of generally accepted accounting procedures allegation)
  2. SEC-11 (allegations of filing false claim with SEC)
  3. INST_INV (whether lead plaintiff is an institution or individuals)
  4. RESTATED_FIN (whether restated financials are involved)
  5. BANKRUPTCY (case is related to bankruptcy filing)
  6. ONE_DAY_DROP (discretized into 2 states: $\leq$ 40.5% and $>$ 40.5%)

Model and Results: What are the parameters?

Naïve Bayes Model (estimated from all 1,150 cases):
with priors for *CLOSING*, likelihoods for *GAAP* and *INST_INV*:

Predictors	Class		
	Dismissed	Settled	Odds/LR
	0.52	0.48	1.08
GAAP			
0	0.80	0.65	1.23
1	0.20	0.35	0.57
INST_INV			
0	0.64	0.51	1.26
1	0.36	0.49	0.73
SEC_11			
0			
1			
RESTATED_FIN			
0			
1			
BANKRUPTCY			
0			
1			
ONE_DAY_DROP			
≤ 0.405 (0)			
> 0.405 (1)			

Model and Results: What are the parameters?

Adding likelihoods for *SEC* – 11:

Predictors	Class: Dismissed	Class: Settled	Odds/LR
	0.52	0.48	1.08
GAAP			
0	0.80	0.65	1.23
1	0.20	0.35	0.57
INST_INV			
0	0.64	0.51	1.26
1	0.36	0.49	0.73
SEC_11			
0	0.70	0.62	1.12
1	0.30	0.38	0.80
RESTATED_FIN			
0			
1			
BANKRUPTCY			
0			
1			
ONE_DAY_DROP			
≤ 0.405 (0)			
> 0.405 (1)			

Rows: SEC_11 Columns: CLOSING

	Dismissed	Settled	All
0	414	346	760
	70%	62%	66%
1	179	211	390
	30%	38%	34%
All	593	557	1150

Model and Results: What are the parameters?

Adding likelihoods for *RESTATED_FIN*:

Predictors	Class: Dismissed	Class: Settled	Odds/LR
	0.52	0.48	1.08
GAAP			
0	0.80	0.65	1.23
1	0.20	0.35	0.57
INST_INV			
0	0.64	0.51	1.26
1	0.36	0.49	0.74
SEC_11			
0	0.70	0.62	1.12
1	0.30	0.38	0.80
RESTATED_FIN			
0	0.93	0.84	1.10
1	0.07	0.16	0.47
BANKRUPTCY			
0			
1			
ONE_DAY_DROP			
≤ 0.405 (0)			
> 0.405 (1)			

Rows: RESTATED_FIN	Columns: CLOSING		
	Dismissed	Settled	All
0	549	469	1018
	93%	84%	89%
1	44	88	132
	7%	16%	11%
All	593	557	1150

Model and Results: What are the parameters?

Adding likelihoods for *BANKRUPTCY*:

Predictors	Class: Dismissed	Class: Settled	Odds/LR
	0.52	0.48	1.08
GAAP			
0	0.80	0.65	1.23
1	0.20	0.35	0.57
INST_INV			
0	0.64	0.51	1.26
1	0.36	0.49	0.74
SEC_11			
0	0.70	0.62	1.26
1	0.30	0.38	0.74
RESTATED_FIN			
0	0.93	0.84	1.10
1	0.07	0.16	0.48
BANKRUPTCY			
0	0.99	0.97	1.03
1	0.01	0.03	0.21
ONE_DAY_DROP			
≤ 0.405 (0)			
> 0.405 (1)			

Rows: BANKRUPTCY	Columns: CLOSING		
	Dismissed	Settled	All
0	589	539	1128
	99%	97%	98%
1	4	18	22
	1%	3%	2%
All	593	557	1150

Model and Results: What are the parameters?

Adding likelihoods for *ONE_DAY_DROP*:

Predictors	Class: Dismissed	Class: Settled	Odds/LR
	0.52	0.48	1.08
GAAP			
0	0.80	0.65	1.23
1	0.20	0.35	0.57
INST_INV			
0	0.64	0.51	1.26
1	0.36	0.49	0.74
SEC_11			
0	0.70	0.62	1.12
1	0.30	0.38	0.80
RESTATED_FIN			
0	0.93	0.84	1.10
1	0.07	0.16	0.48
BANKRUPTCY			
0	0.99	0.97	1.03
1	0.01	0.03	0.21
ONE_DAY_DROP			
≤ 0.405 (0)	0.89	0.85	1.05
> 0.405 (1)	0.11	0.15	0.74

Rows: One_Day_Drop Columns: CLOSING

	Dismissed	Settled	All
0	230	208	438
	89%	85%	87%
1	28	36	64
	11%	15%	13%
Non-Missing	258	244	502
Missing	335	313	648
All	593	557	1150

Model and Results: What are the parameters?

Model:

Evidence	Class		Odds/ Likelihood Ratio	Comment
	Dismissed	Settled		
Prior	0.52	0.48	1.08	
GAAP Violation? No	0.80	0.65	1.23	Favors *dismissed*
GAAP Violation? Yes	0.20	0.35	0.57	Favors *settled*
Lead Plaintiff Institution? No	0.64	0.51	1.26	Favors *dismissed*
Lead Plaintiff Institution? Yes	0.36	0.49	0.74	Favors *settled*
SEC False Filing? No	0.70	0.62	1.12	Favors *dismissed*
SEC False Filing? Yes	0.30	0.38	0.80	Favors *settled*
Re-stated Financials? No	0.93	0.84	1.10	Favors *dismissed*
Re-stated Financials? Yes	0.07	0.16	0.48	Favors *settled*
Bankruptcy? No	0.99	0.97	1.03	Favors *dismissed*
Bankruptcy? Yes	0.01	0.03	0.21	Favors *settled*
Largest 1-day Drop is ≤ 40.5%	0.89	0.85	1.05	Favors *dismissed*
Largest 1-day Drop is > 40.5%	0.11	0.15	0.74	Favors *settled*

Model and Results: Using our NB model

- If all predictors are No:

Evidence		Odds/Likelihood Ratios for *dismissed*
Prior		1.08
GAAP Violation?	No	1.23
SEC False Filing?	No	1.12
Lead Plaintiff Institution?	No	1.26
Re-stated Financials?	No	1.10
Bankruptcy?	No	1.03
Largest 1-day Drop in Stock Price	≤ 40.5%	1.05

  - Odds for *dismissed* $= 1.08 \cdot 1.23 \cdot 1.12 \cdot 1.26 \cdot 1.10 \cdot 1.03 \cdot 1.05 = 2.22$
  - Probability of *dismissed* $= \frac{2.22}{3.22} = 0.69$

Model and Results: Using our NB model

- If all predictors are Yes:

Evidence		Odds/Likelihood Ratios for *dismissed*
Prior		1.08
GAAP Violation?	Yes	0.57
SEC False Filing?	Yes	0.80
Lead Plaintiff Institution?	Yes	0.74
Re-stated Financials?	Yes	0.48
Bankruptcy?	Yes	0.21
Largest 1-day Drop in Stock Price	≥ 40.5%	0.74

  - Odds for *dismissed* $= 1.08 \cdot 0.57 \cdot 0.80 \cdot 0.74 \cdot 0.48 \cdot 0.21 \cdot 0.74 = 0.03$
  - Probability of *dismissed* $= \frac{0.03}{1.03} = 0.03$

Model and Results: Using our NB model

- If predictors are as follows (e.g., Panera Bread Company, 2008, E. D. Missouri):

Evidence		Odds/Likelihood Ratios for *dismissed*
Prior		1.08
GAAP Violation?	Yes	0.57
SEC False Filing?	No	1.12
Lead Plaintiff Institution?	Yes	0.74
Re-stated Financials?	No	1.26
Bankruptcy?	No	1.03
Largest 1-day Drop in Stock Price	?	

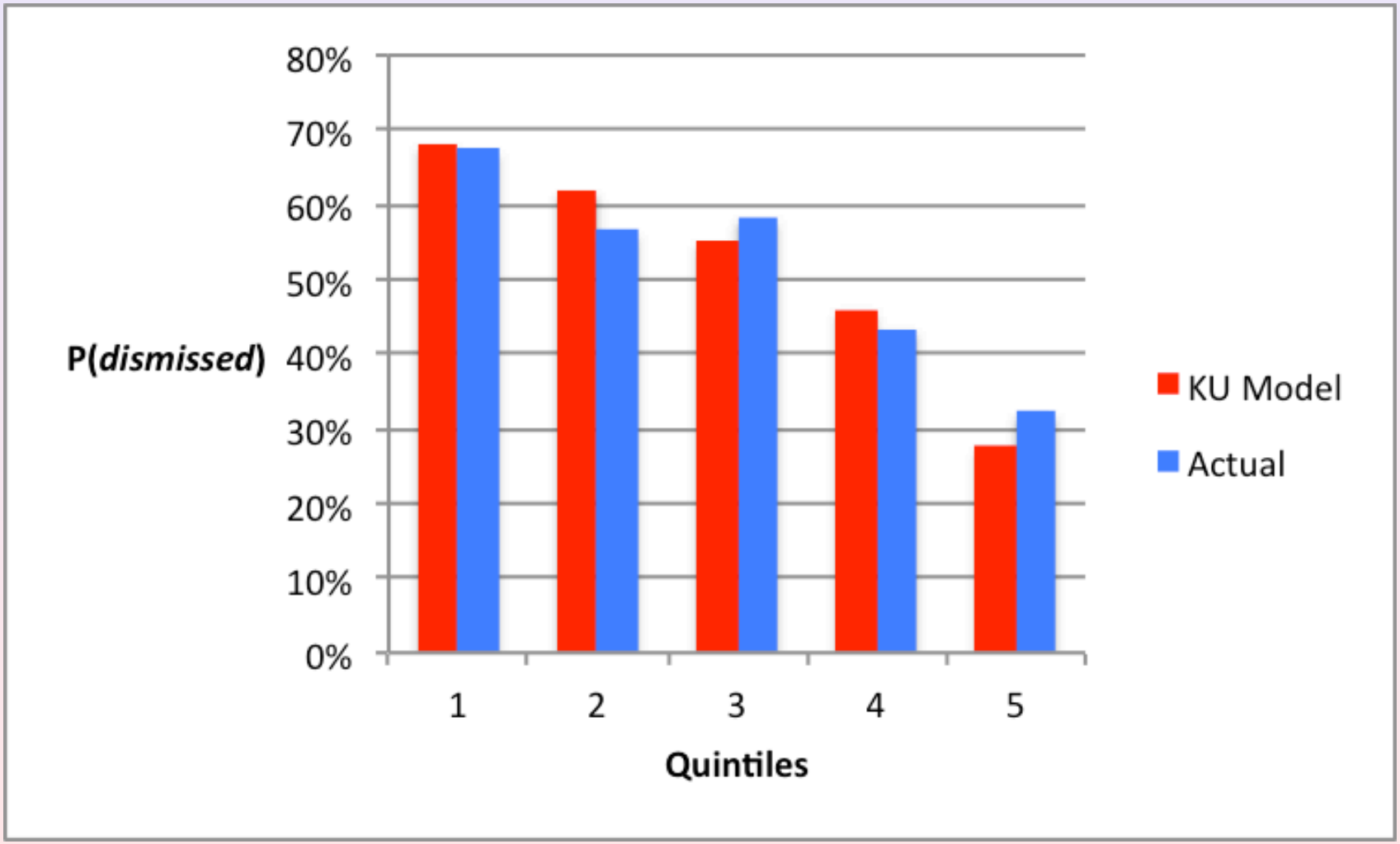
  - Odds for *dismissed* $= 1.08 \cdot 0.57 \cdot 1.12 \cdot 0.74 \cdot 1.26 \cdot 1.03 = 0.58$
  - Probability of *dismissed* $= \frac{0.58}{1.58} = 0.37$

Model and Results: How good is our model?

- For each case, we compute the *Pr*(*dismissed*) based on a naïve Bayes model whose parameters are estimated from the other 1,149 cases (Lachenbruch procedure).
- We sort the cases by *Pr*(*dismissed*), highest to lowest, and divide the set of all cases into 5 groups of 230 cases each (quintiles). The characteristics of each quintile are as follows:

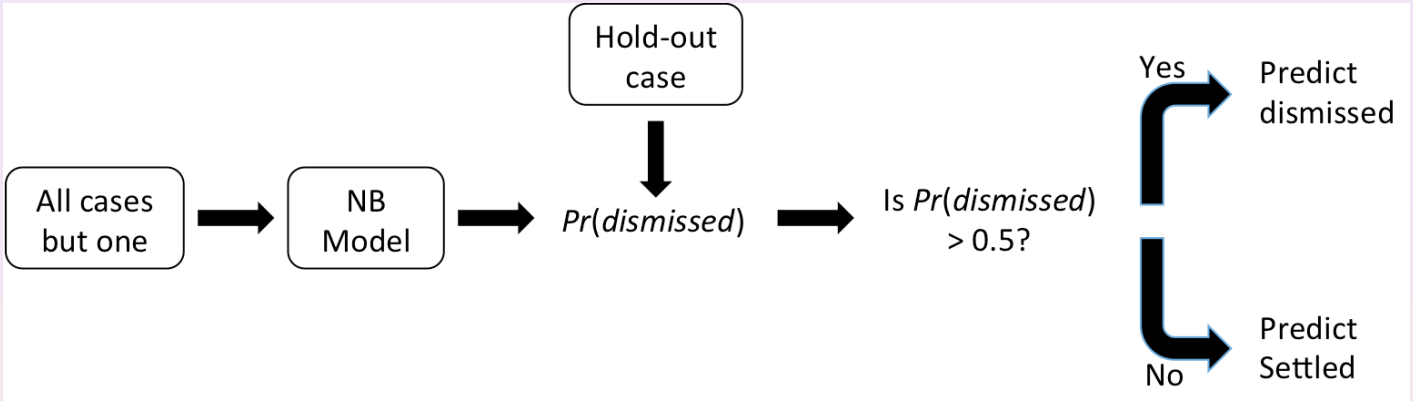
Using Lachenbruch	Quintile					
	# 1	# 2	# 3	# 4	# 5	All
Max Pr(dismissed)	69%	68%	60%	50%	37%	69%
Min Pr(Dismissed)	68%	60%	50%	37%	7%	7%
Avg. Pr(Dismissed)	**68%**	**62%**	**55%**	**46%**	**28%**	**52%**
# dismissed	131	150	135	101	76	593
# settled	99	80	95	129	154	557
Proportion dismissed	**57%**	**65%**	**59%**	**44%**	**33%**	**52%**
Proportion settled	43%	35%	41%	56%	67%	48%

Model and Results: How good is our model?



Model and Results: How good is our model?

How well can our NB model predict *CLOSING*?:



- This procedure is repeated with each case as a hold-out case (Lachenbruch)

Model and Results: How good is our model?

- Confusion matrix using the Lachenbruch procedure:

# cases		Predicted		
		Dismissed	Settled	Totals
Actual	Dismissed	423	170	593
	Settled	274	283	557

- # incorrect predictions is $274 + 170 = 444$ cases (39%)
- A naïve strategy of predicting all cases as dismissed would have an error of 557 (48%)
- So "lift over marginal" is $557 - 444 = 113$ cases (10%)
- The search method for identifying a good subset uses # incorrect predictions as an objective to be minimized

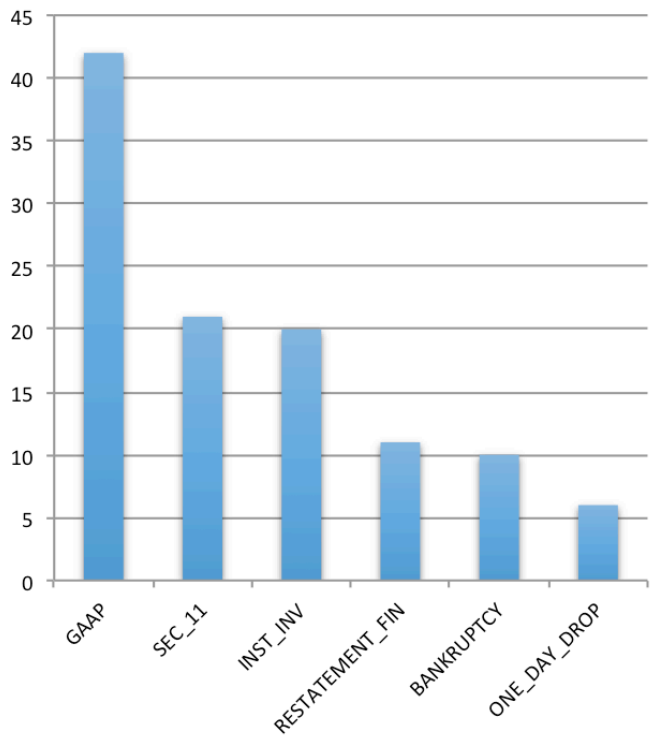
Model and Results: Relative influence of each predictor

What is the relative influence of each predictor variable?

- For each variable, we removed it from the set of 6 predictor variables, and observed the increase in *# incorrect predictions* as a result.
- Results are as follows:

Relative Influence of Variables:

Variable	Increase in # incorrect predictions (cases)	% of Total
GAAP	42	38%
SEC_11	21	19%
INST_INV	20	18%
RESTATEMENT_FIN	11	10%
BANKRUPTCY	10	9%
ONE_DAY_DROP	6	5%
TOTAL	110	100%



Model and Results: Relative influence of each predictor

- Relative influence of predictors depends on likelihood ratio, and frequency of occurrence:
  - Smaller the likelihood ratio, higher the influence
  - Higher the frequency of occurrence, higher the influence

	Likelihood Ratio for *dismissed*	Frequency, cases	Increase in prediction error, cases
GAAP Violation is yes	0.57	310	42
SEC False Filing is yes	0.80	390	21
Lead Plaintiff Institution is yes	0.74	490	20
Re-stated Financials is yes	0.48	132	11
Bankruptcy is yes	0.21	22	10
Largest 1-day drop is > 40.5%	0.74	64	6

Outline

- Problem
- Data
- Model and Results
- Limitations
- Future Work

Limitations

- History of past cases is used to predict the future, which assumes that future will be like the past
- Changes in accounting rules or litigation laws could change the filing and closing of securities class action cases
- We have too many (56%) missing values for ONE_DAY_DROP. Our model may perform better if we had fewer missing values
- Model built is only as good as the data—errors in the data will result in errors in the model

Outline

- Problem
- Data
- Model and Results
- Limitations
- Future Work

Future Work

- For each case, we know the judge who made the judgment to dismiss/not dismiss. Judges have a history of ruling on civil cases. Also we know which president (Democrat or Republican) appointed the judges. Can we use such information to improve our model?
- Financial re-statements can be classified as *core*/*non-core*, materially significant or not. Can we take advantage of such classifications of financial re-statements to predict closing?

Future Work

- Are there other variables (not in Advisen data set) that we can use to predict closing? E.g.,
  - Stock options for CEO and board members that provide perverse incentives to ratchet up stock prices at exercise dates (C Shenoy)
  - Conservatism of reporting earnings (Ettredge)
  - Insider selling of securities based on insider information
  - Short interest (Meschke)
  - Notoriety of the class action case (Meschke)