Aspects of
Model Performance

School of Information Studies
Syracuse University

Aspects of Model Performance

Accuracy: total correct predictions/total

Speed

- Time to construct model (training time)
- Time to use the model (prediction time)

Robustness: handling noise and missing values

Scalability: efficiency in handling large data set

Interpretability

- Understanding and insight provided by the model

School of Information Studies
Syracuse University

Compare DT and MNB Time

Time for training model

Time for prediction

```
Size of the tree :        411

Time taken to build model: 17.5 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.17 seconds


Time taken to build model: 1.13 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.16 seconds
```

School of Information Studies
Syracuse University

MNB Robustness

Robustness: handling noise and missing values

- Noise?
  - If you change an example's label from "pos" to "neg," how would that affect the model's performance?

- Missing value?
  - Does it exist in text vectors?

School of Information Studies
Syracuse University

MNB Scalability

Scalability: efficiency in handling large data set

Can the probabilities be calculated using parallel processing?

School of Information Studies
Syracuse University

MNB Interpretability

MNB is a linear model:

$$\hat{P}(c|d) \propto \hat{P}(c) \prod_{1 \le k \le n_d} \hat{P}(t_k|c)$$

$$\log \frac{\hat{P}(c|d)}{\hat{P}(\bar{c}|d)} = \log \frac{\hat{P}(c)}{\hat{P}(\bar{c})} + \sum_{1 \le k \le n_d} \log \frac{\hat{P}(t_k|c)}{\hat{P}(t_k|\bar{c})}$$

Feature weight

odds ratio

School of Information Studies
Syracuse University