

Specific Subcategories May be Noisy

From the data, for example,

$$\text{prob}(\text{N AND I AND F} \mid \text{R}) \times \text{prob}(\text{R}) = \\ (4/27) \times (27/200) = 0.0200$$

$$\text{prob}(\text{N AND I AND F} \mid \text{B}) \times \text{prob}(\text{B}) = \\ (3/27) \times (27/200) = 0.0150$$

$$\text{prob}(\text{N AND I AND F} \mid \text{W}) \times \text{prob}(\text{W}) = \\ (4/83) \times (83/200) = 0.0200$$

$$\text{prob}(\text{N AND I AND F} \mid \text{G}) \times \text{prob}(\text{G}) = \\ (2/64) \times (63/200) = 0.0100$$

Blue is the best guess for the model used to generate the data. These are both incorrect due to statistical variation.