In general, the procedures for me to find out good features could be divided into three steps as the following:

1. To begin with, I firstly rationally think of some factors that should influence the correctness of the guesses. For example, I believe if the current shown contents of a question (run) contains too many noninformative words such as ‘a’, ‘the’ or ‘of’, then it is hard to guess correctly, so I let the number of these useless words to be a feature; The year in which the question was asked can reflect how difficult the question was and influence the correctness, and similarly the ‘Difficulty’ and ‘Tournament’ attributes of the question reflects similar information, so I add all of them as features; The ‘Category’ of a question might matter because sometimes it is just harder to guess some genre of questions than the other.
2. After adding these ‘reasonable’ features, I begin to test individually that if having one single feature can I beat the performance of that with no feature.

The baseline is no feature model trained with 500 training samples and its performance on 50 test samples. The result is 167 right out of 407 with Accuracy: 0.73 Buzz ratio: 0.32 Buzz position: -0.073743 Best rate: 0.41.I focused on the right number and best rate.

‘Difficulty’ gives me best rate of 0.42 and right number as 172 which improves a little bit.