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 the length of run should be a different feature from the length of guess because with shorter part of the question it will be harder to guess, so I also include 'LengthRun'; 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 called 'UselessInfo'; Similarly, I count the number of words indicating genders like 'he, she, it, they, his, her, its' as a feature 'GenderInfo' because for question on people's name these words would help a lot; 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.

   'UselessInfo' gives me best rate of 0.42 and right number as 169 which improves a little bit. The more important point is that I find this feature does vary among different questions as shown in the screenshot below.

```
================
best 0.42
==================

              guess: Kent State shootings
             answer: Kent_State_shootings
                 id: 93193
     gpr_confidence: -0.1176
  UselessInfo_guess: 10
               text: A letter written after this event by one-time Alaskan Independence
                     Party leader Wally Hickel led to Hickel's firing as Secretary of the
                     Interior. A 2010 forensic study into the Strubbe tape implied that
                     Terry Norman provoked this event, which helped inspire the unrest of
                     New York's Hard Hat Riot four days later. The Scranton Commission
                     investigated this event. Governor Jim Rhodes threatened martial law in
                     the run-up to this event, whose aftermath was shown in John Filo's
                     photograph of a kneeling woman screaming. This event happened near
                     Blanket Hill in the Prentice Hall parking lot, as its victims
                     protested Nixon's decision to expand the Vietnam War into Cambodia.
                     For ten points, name this 1970
--------------------
              guess: Red Sea
             answer: Red_Sea
                 id: 93167
     gpr_confidence: -0.0052
  UselessInfo_guess: 4
               text: This geographic feature was closed to Christians by traders called
                     Karimi after Reynaud of Chatillon irked them. Purported cave dwellers
                     on this body of water's western side were the first people called
                     "Troglodytes." A port called "Mussel Harbor" abutted this body near
                     Berenice according to an anonymous 1st-century text about its peoples.
                     The city of Adulis traded with the Himyarite kingdom across
```

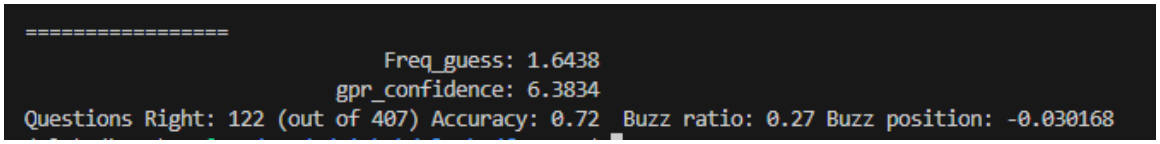
The first guess has 10 useless words and the second has 4.

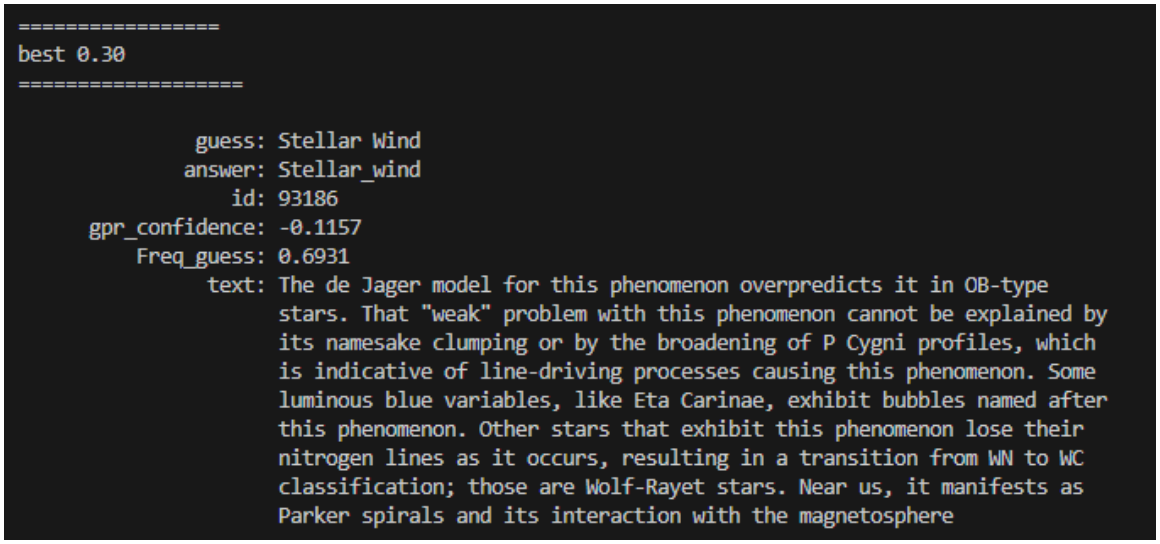
On the opposite, among the 50 test samples the features ‘Year’ are almost the same as shown below.

```
--------------------
              guess: Vulture
             answer: Vultures
                 id: 93141
     gpr_confidence: -0.1129
         Year_guess: 2015
               text: Some Vajrayana Buddhists consider these real-world creatures to be
                     Dakini, a type of angelic psychopomp. They are propitiated at
                     buildings made of three concentric stone circles of varying height. In
                     a ritual meant to satisfy these creatures, a master known as a rogyapa
                     uses a slicing knife during readings from the Tibetan Book of the
                     Dead. On a peak named for these creatures near Ramnagar, the Heart
                     Sutra and Lotus Sutra were delivered by the Buddha. When not shown as
                     an eagle, Garuda's brother Jatayu is one of these creatures, whose
                     recent chemical-caused extinction around Mumbai has threatened the use
                     of dakhmas there by Parsis. For 10 points, name these birds which come
                     to Tibetan "sky-burials"
--------------------
              guess: Leonhard Euler
             answer: Peter_Gustav_Lejeune_Dirichlet
                 id: 93240
     gpr_confidence: -0.0118
         Year_guess: 2015
               text: The stick-breaking process is a constructive algorithm used to
                     generate samples from a stochastic process named after this
                     mathematician. This mathematician also names a multivariate continuous
                     probability distribution defined on a n-dimensional simplex, which is
                     the conjugate prior of the multinomial distribution. In analytic
                     number theory, the Riemann zeta function is a simple non-trivial
                     example of his type of series, which are sums over n of a-sub-n over n
                     to a complex power s. This mathematician also names the eta function,
                     a simple example of one of his L-functions. In contrast to Neumann
                     boundary conditions, boundary conditions named after this man specify
                     a function's value on the boundary of its domain. For 10 points, name
                     this man who codified, and sometimes names, the pigeonhole principle.
--------------------
              guess: Narcissistic personality disorder
             answer: Narcissism
                 id: 93168
     gpr_confidence: -0.1198
         Year_guess: 2015
               text: The nature of this condition was debated by Heinz Kohut and Otto
                     Kernberg. In an essay on this condition,
```

The Best rate and right number are the same as the baseline. The comparison proves my hypothesis that if the features are almost the same for most of the samples, then it cannot really help the classifier.

With such rule, I delete features that don't vary a lot among different samples such as 'Year', 'Prompt', 'GamePlay' and 'Tournament'. I also delete 'Frequency' because every time it brings all indices down as shown below:





I tried out the combination of 'Frequency' and other features but still cannot save these indices, so I finally delete this.

I remain 'Difficulty', 'Length', 'LengthRun', 'UselessInfo' and 'GenderInfo' because these features are different among different guesses and bring me good indices locally. By the way I thought encoding 'Difficulty' into integer might help the classifier better, but by test I find the performance is better if remaining the feature to be in string type.

3. Since I already filtered the pool of features, I can try some combinations of picked features on Gradescope and see their performances. Of course, to include all the features will cause overfit so I only use some subsets. Sometimes a combination will have very high accuracy like 0.774, but medium Best Score 0.414. As the instruction said the Best Score is more important than Acc, then I finally choose a combination with balanced score, which is ['UsefulInfo','Length','GenderInfo','LengthRun']

4	Minghan Yu	0.7703730315549967	0.4307526495419436	0.3470151487934854

I beat the baseline a lot and am at a good ranking until the time point when I wrote this analysis. In all I used some reasonable features based on human judgment and it works somehow.