Problem 1

Since $\epsilon \sim N(0, \sigma^2)$, $y|x, \beta \sim N(x^T\beta, \sigma^2)$,

$$\begin{aligned}
L &= \prod_{i=1}^{n} P(y_i|x_i, \beta) \\
&= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\sum_{i=1}^{n} \frac{(y - x^T\beta)^2}{2\sigma^2}\right) \\
\log L &= -\sum_{i=1}^{n} \frac{(y - x^T\beta)^2}{2\sigma^2} - n\log(\sqrt{2\pi}\sigma) \\
AIC &= -2\log L + 2d \\
&= \frac{RSS}{\sigma^2} + 2d + 2n\log(\sqrt{2\pi}\sigma)
\end{aligned}$$

Noticed that $2n\log(\sqrt{2\pi}\sigma)$ is constant for certain data, we conclude $AIC = \dfrac{RSS}{\sigma^2} + 2d$.

Problem 2

There are two main problems in this approach.

(1).There were too many garbage features in the model. For example, for MI, they had over 24,000 features, most of which had nothing to do with the output. As a result, the model's resources were wasted and performance suffered.

(2).The way they labeled data was unreasonable. Y depended on only one predictor. Although after L1-penalized logistic regression there might be more than one active features, the performance of the model would degrade to near single-variable model.