

























Data Quality  
Data Cleaning  
Noise























































































SHOW ME YOUR FRIENDS  
I WILL TELL YOU WHO  
YOU ARE





























































$p$  is the predicted probability,  
 $\beta_0$  is the intercept,  
 $\beta_1, \beta_2, \dots, \beta_n$  are the coefficients for the input features  $X_1, X_2, \dots, X_n$ ,  
 $e$  is the base of the natural logarithm.









































**THANK YOU**