* q background?

Hi, my name is Dave. I am currently working for shell oil company in Houston, Texas as a data scientist.

I have been in the field of data science for more than 5 years since being working for Axiom Tech Group,

I am very familiar with data pipeline building and deployment of different machine learning and neural network using a variety of systems, these parts are what i am in charge of. My industry experience includes predictive analytics in consulting, marketing, finance, logistic area.

My project is wrapping up, about to hand over to devs op team. we made good progress. so i'm seeking for a new opportunity.

I hope my solid background and experience can contribute to your success.

* q example of a project you are working on?

The Predictive Analytics Project I'm working on is supposed to enhance oil field production

and cuts costs by finding optimal well settings and forecasting equipment failures and potential problems.

* q framework for ML?

I used tensorflow, pytorch, keras, scikit-learn framework, I also used sparkmil to implyment machine learning on the cloud.

* q what is a bias?

In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa.

* q simple model you tried for your project?

I usually work with logistic regression model, this is probably the simplest model but it is explainable and often perform very well.

* q what is a ROC

ROC curve is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True positive rate, False positive rate.

[AUC](https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc) is an area, AUC provides an aggregate measure of performance across all possible classification thresholds.

* q KNN vs Kmeans?

[K-means](http://en.wikipedia.org/wiki/K_means) is a clustering algorithm that tries to partition a set of points into K sets (clusters) such that the points in each cluster tend to be near each other. It is unsupervised because the points have no external classification.

[K-nearest neighbors](http://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm) is a classification (or regression) algorithm that in order to determine the classification of a point, combines the classification of the K nearest points. It is supervised because you are trying to classify a point based on the known classification of other points.

* q what is a slope?

it represents how much y changes as to increase or decrease of x.

* q role at the project?

I worked as a data scientist at the project.

* q what is in the data?

We had text data, time series data to deal with. I even had some climate data.

* q why LSTM on your last project?

We were we doing text classification project, LSTM makes it easier for inputs to be repeated without much alteration. This does not sounds really interesting but in fact enables important information to be kept for a long time: it does not forget information just because of time (note that he can actively forget tho).

* q why are you looking for a change?

My project is wrapping up, about to hand over to devs op team. we made good progress. so i'm seeking for a new opportunity.

I hope my solid background and experience can contribute to your success.