**Analytics & Data Science Questions**

**Q1.** **Explain machine learning to me like a 5 year old.**

**A1.** It’s simple. It’s just like how babies learn to walk. Every time they fall down, they learn (unconsciously) & realize that their legs should be straight and not in a bend position. The next time they fall down, they feel pain. They cry. But, they learn ‘not to stand like that again’. In order to avoid that pain, they try harder. To succeed, they even seek support from the door or wall or anything near them, which helps them stand firm.

This is how a machine works & develops intuition from its environment.

Broadly speaking these are the steps. Of course these may vary slightly depending on the type of problem, data, tools available etc.

1. Problem definition – The first step is to of course understand the business problem. What is the problem you are trying to solve – what is the business context? Very often however your client may also just give you a whole lot of data and ask you to do something with it. In such a case you would need to take a more exploratory look at the data. Nevertheless if the client has a specific problem that needs to be tackled, then then first step is to clearly define and understand the problem. Yo2u will then need to convert the business problem into an analytics problem. I other words you need to understand exactly what you are going to predict with the model you build. There is no point in building a fabulous model, only to realise later that what it is predicting is not exactly what the business needs.

2. Data Exploration – Once you have the problem defined, the next step is to explore the data and become more familiar with it. This is especially important when dealing with a completely new data set.

3. Data Preparation – Now that you4 have a good understanding of the data, you will need to prepare it for modelling. You will identify and treat missing values, detect outliers, transform variables, create binary variables if required and so on. This stage is very influenced by the modelling technique you will use at the next stage.  For example, regression involves a fair amount of data preparation, but decision trees may need less prep whereas clustering requires a whole different kind of prep as compared to other techniques.

4. Modelling – Once the data is prepared, you can begin modelling. This is usually an iterative process where you run a model, evaluate the results, tweak your approach, run another model, evaluate the results, re-tweak and so on….. You go on doing this until you come up with a model you are satisfied with or what you feel is the best possible result with the given data.

5. Validation – The final model (or maybe the best 2-3 models) should then be put through the validation process. In this process, you test the model using completely new data set i.e. data that was not used to build the model. This process ensures that your model is a good model in general and not just a very good model for the specific data earlier used (Technically, this is called avoiding over fitting)

6. Implementation and tracking – The final model is chosen after the validation. Then you start implementing the model and tracking the results. You need to track results to see the performance of the model over time. In general, the accuracy of a model goes down over time. How much time will really depend on the variables – how dynamic or static they are, and the general environment – how static or dynamic that is.

Q 2. What do you do in data exploration?

A 2. Data exploration is done to become familiar with the data. This step is especially important when dealing with new data. There are a number of things you will want to do in this step –

a.        What is there in the data – look at the list of all the variables in the data set. Understand the meaning of each variable using the data dictiona2ry. Go back to the business for more information in case of any confusion.

b.        How much data is there – look at the volume of the data (how many records), look at the time frame of the data (last 3 months, last 6 months etc.)

c.         Quality of the data – how much missing information, quality of data in each variable. Are all fields usable? If a field has data for only 10% of the observations, then maybe that field is not usable etc.

d.        You will also identify some important variables and may do a deeper investigation of these. Like looking at averages, min and max values, maybe 10th and 90th percentile as well…

e.        You may also identify fields that you need to transform in the data prep stage.

In data preparation, you will prepare the data for the next stage i.e. the modelling stage. What you do here is influenced by the choice of technique you use in the next stage.

But some things are done in most cases – example identifying missing values and treating them, identifying outlier values (unusual values) and treating them, transforming variables, creating binary variables if required etc,

This is the stage where you will partition the data as well. i.e create training data (to do modelling) and validation (to do validation).

 Q3: How will you treat missing values?

A3. The first step is to identify variables with missing values. Assess the extent of missing values. Is there a pattern in missing values? If yes, try and identify the pattern. It may lead to interesting insights.

If no pattern, then we can either ignore missing values (SAS will not use any observation with missing data) or impute the missing values.

Q4. How will you treat outlier values?

A4. You can identify outliers using graphical analysis and univariate analysis. If there are only a few outliers, you can assess them individually. If there are many, you may want to substitute the outlier values with the 1stpercentile or the 99th percentile values.

If there is a lot of data, you may decide to ignore records with outliers.

Not all extreme values are outliers. Not all outliers are extreme values.

Q5. List out some of the best practices for data cleaning?

A5. Some of the best practices for data cleaning includes,

* Sort data by different attributes
* For large datasets cleanse it stepwise and improve the data with each step until you achieve a good data quality
* For large datasets, break them into small data. Working with less data will increase your iteration speed
* To handle common cleansing task create a set of utility functions/tools/scripts. It might include, remapping values based on a CSV file or SQL database or, regex search-and-replace, blanking out all values that don’t match a regex
* If you have an issue with data cleanliness, arrange them by estimated frequency and attack the most common problems
* Analyze the summary statistics for each column ( standard deviation, mean, number of missing values,)
* Keep track of every date cleaning operation, so you can alter changes or remove operations if required

Q6. List out some common problems faced while analyzing data?

A6. Some of the common problems faced by data analyst are

* Common misspelling
* Duplicate entries
* Missing values
* Illegal values
* Varying value representations
* Identifying overlapping data

Q7. Explain what should be done with suspected or missing data?

A7.

* Prepare a validation report that gives information of all suspected data. It should give information like validation criteria that it failed and the date and time of occurrence
* Experience personnel should examine the suspicious data to determine their acceptability
* Invalid data should be assigned and replaced with a validation code
* To work on missing data use the best analysis strategy like deletion method, single imputation methods, model based methods, etc.

Q8. Mention what are the key skills required for a Data Scientist?

A8. A data scientist must have the following skills

* Database knowledge
* Database management
* Data blending
* Querying
* Data manipulation
* Predictive Analytics
* Basic descriptive statistics
* Predictive modeling
* Advanced analytics
* Big Data Knowledge
* Big data analytics
* Unstructured data analysis
* Machine learning
* Presentation skill
* Data visualization
* Insight presentation
* Report design

**Q9.** **You are given a data set. The data set has missing values which spread along 1 standard deviation from the median. What percentage of data would remain unaffected? Why?**

**A9.** This question has enough hints for you to start thinking! Since, the data is spread across median, let’s assume it’s a normal distribution. We know, in a normal distribution, ~68% of the data lies in 1 standard deviation from mean (or mode, median), which leaves ~32% of the data unaffected. Therefore, ~32% of the data would remain unaffected by missing values.

Q10. You are given a data set on cancer detection. You’ve build a classification model and achieved an accuracy of 96%. Why shouldn’t you be happy with your model performance? What can you do about it?

A10. If you have worked on enough data sets, you should deduce that cancer detection results in imbalanced data. In an imbalanced data set, accuracy should not be used as a measure of performance because 96% (as given) might only be predicting majority class correctly, but our class of interest is minority class (4%) which is the people who actually got diagnosed with cancer. Hence, in order to evaluate model performance, we should use Sensitivity (True Positive Rate), Specificity (True Negative Rate), F measure to determine class wise performance of the classifier. If the minority class performance is found to be poor, we can undertake the following steps:

1. We can use undersampling, oversampling or SMOTE to make the data balanced.
2. We can alter the prediction threshold value by doing [probability caliberation](https://www.analyticsvidhya.com/blog/2016/07/platt-scaling-isotonic-regression-minimize-logloss-error/) and finding a optimal threshold using AUC-ROC curve.
3. We can assign weight to classes such that the minority classes gets larger weight.
4. We can also use anomaly detection.

Q11. Explain prior probability, likelihood and marginal likelihood in context of naiveBayes algorithm?

A11.  Prior probability is nothing but, the proportion of dependent (binary) variable in the data set. It is the closest guess you can make about a class, without any further information. For example: In a data set, the dependent variable is binary (1 and 0). The proportion of 1 (spam) is 70% and 0 (not spam) is 30%. Hence, we can estimate that there are 70% chances that any new email would  be classified as spam.

Likelihood is the probability of classifying a given observation as 1 in presence of some other variable. For example: The probability that the word ‘FREE’ is used in previous spam message is likelihood. Marginal likelihood is, the probability that the word ‘FREE’ is used in any message.

**Q12.** **You are working on a time series data set. You manager has asked you to build a high accuracy model. You start with the decision tree algorithm, since you know it works fairly well on all kinds of data. Later, you tried a time series regression model and got higher accuracy than decision tree model. Can this happen? Why?**

**A12.** Time series data is known to posses linearity. On the other hand, a decision tree algorithm is known to work best to detect non – linear interactions. The reason why decision tree failed to provide robust predictions because it couldn’t map the linear relationship as good as a regression model did. Therefore, we learned that, a linear regression model can provide robust prediction given the data set satisfies its [linearity assumptions](https://www.analyticsvidhya.com/blog/2016/07/deeper-regression-analysis-assumptions-plots-solutions/).

Q13. You came to know that your model is suffering from low bias and high variance. Which algorithm should you use to tackle it? Why?

A13.  Low bias occurs when the model’s predicted values are near to actual values. In other words, the model becomes flexible enough to mimic the training data distribution. While it sounds like great achievement, but not to forget, a flexible model has no generalization capabilities. It means, when this model is tested on an unseen data, it gives disappointing results.

In such situations, we can use bagging algorithm (like random forest) to tackle high variance problem. Bagging algorithms divides a data set into subsets made with repeated randomized sampling. Then, these samples are used to generate  a set of models using a single learning algorithm. Later, the model predictions are combined using voting (classification) or averaging (regression).

Also, to combat high variance, we can:

1. Use regularization technique, where higher model coefficients get penalized, hence lowering model complexity.
2. Use top n features from variable importance chart. May be, with all the variable in the data set, the algorithm is having difficulty in finding the meaningful signal.

**Q14.** **How is kNN different from kmeans clustering?**

**A14.**  Don’t get mislead by ‘k’ in their names. You should know that the fundamental difference between both these algorithms is, kmeans is unsupervised in nature and kNN is supervised in nature. kmeans is a clustering algorithm. kNN is a classification (or regression) algorithm.

kmeans algorithm partitions a data set into clusters such that a cluster formed is homogeneous and the points in each cluster are close to each other. The algorithm tries to maintain enough separability between these clusters. Due to unsupervised nature, the clusters have no labels.

kNN algorithm tries to classify an unlabeled observation based on its k (can be any number ) surrounding neighbors. It is also known as lazy learner because it involves minimal training of model. Hence, it doesn’t use training data to make generalization on unseen data set.

**Q15.** **After analyzing the model, your manager has informed that your regression model is suffering from multicollinearity. How would you check if he’s true? Without losing any information, can you still build a better model?**

**A15.** To check multicollinearity, we can create a correlation matrix to identify & remove variables having correlation above 75% (deciding a threshold is subjective). In addition, we can use calculate VIF (variance inflation factor) to check the presence of multicollinearity. VIF value <= 4 suggests no multicollinearity whereas a value of >= 10 implies serious multicollinearity. Also, we can use tolerance as an indicator of multicollinearity.

But, removing correlated variables might lead to loss of information. In order to retain those variables, we can use penalized regression models like ridge or lasso regression. Also, we can add some random noise in correlated variable so that the variables become different from each other. But, adding noise might affect the prediction accuracy, hence this approach should be carefully used.

**Q16. When is Ridge regression favorable over Lasso regression?**

**A16.** You can quote ISLR’s authors Hastie, Tibshirani who asserted that, in presence of few variables with medium / large sized effect, use lasso regression. In presence of many variables with small / medium sized effect, use ridge regression.

Conceptually, we can say, lasso regression (L1) does both variable selection and parameter shrinkage, whereas Ridge regression only does parameter shrinkage and end up including all the coefficients in the model. In presence of correlated variables, ridge regression might be the preferred choice. Also, ridge regression works best in situations where the least square estimates have higher variance. Therefore, it depends on our model objective.

Q17. While working on a data set, how do you select important variables? Explain your methods.

A17. Following are the methods of variable selection you can use:

1. Remove the correlated variables prior to selecting important variables
2. Use linear regression and select variables based on p values
3. Use Forward Selection, Backward Selection, Stepwise Selection
4. Use Random Forest, Xgboost and plot variable importance chart
5. Use Lasso Regression
6. Measure information gain for the available set of features and select top n features accordingly.

**Q18.** **What is the difference between covariance and correlation?**

**A18.** Correlation is the standardized form of covariance.

Covariances are difficult to compare. For example: if we calculate the covariances of salary ($) and age (years), we’ll get different covariances which can’t be compared because of having unequal scales. To combat such situation, we calculate correlation to get a value between -1 and 1, irrespective of their respective scale.

**Q19. Both being tree based algorithm, how is random forest different from Gradient boosting algorithm (GBM)?**

**A19.** The fundamental difference is, random forest uses bagging technique to make predictions. GBM uses boosting techniques to make predictions.

In bagging technique, a data set is divided into n samples using randomized sampling. Then, using a single learning algorithm a model is build on all samples. Later, the resultant predictions are combined using voting or averaging. Bagging is done is parallel. In boosting, after the first round of predictions, the algorithm weighs misclassified predictions higher, such that they can be corrected in the succeeding round. This sequential process of giving higher weights to misclassified predictions continue until a stopping criterion is reached.

Random forest improves model accuracy by reducing variance (mainly). The trees grown are uncorrelated to maximize the decrease in variance. On the other hand, GBM improves accuracy my reducing both bias and variance in a model.

**Q20.** **You’ve got a data set to work having p (no. of variable) > n (no. of observation). Why is OLS as bad option to work with? Which techniques would be best to use? Why?**

**A20.** In such high dimensional data sets, we can’t use classical regression techniques, since their assumptions tend to fail. When p > n, we can no longer calculate a unique least square coefficient estimate, the variances become infinite, so OLS cannot be used at all.

To combat this situation, we can use penalized regression methods like lasso, LARS, ridge which can shrink the coefficients to reduce variance. Precisely, ridge regression works best in situations where the least square estimates have higher variance.

Among other methods include subset regression, forward stepwise regression.

**Q21.** **We know that one hot encoding increasing the dimensionality of a data set. But, label encoding doesn’t. How ?**

**A21.** Don’t get baffled at this question. It’s a simple question asking the difference between the two.

Using one hot encoding, the dimensionality (a.k.a features) in a data set get increased because it creates a new variable for each level present in categorical variables. For example: let’s say we have a variable ‘color’. The variable has 3 levels namely Red, Blue and Green. One hot encoding ‘color’ variable will generate three new variables as Color.Red, Color.Blue and Color.Green containing 0 and 1 value.

In label encoding, the levels of a categorical variables gets encoded as 0 and 1, so no new variable is created. Label encoding is majorly used for binary variables.

Q22. You are given a data set consisting of variables having more than 30% missing values? Let’s say, out of 50 variables, 8 variables have missing values higher than 30%. How will you deal with them?

A22. We can deal with them in the following ways:

1. Assign a unique category to missing values, who knows the missing values might decipher some trend
2. We can remove them blatantly.
3. Or, we can sensibly check their distribution with the target variable, and if found any pattern we’ll keep those missing values and assign them a new category while removing others.

**Q23.** **‘People who bought this, also bought…’ recommendations seen on amazon is a result of which algorithm?**

**A23.** The basic idea for this kind of recommendation engine comes from collaborative filtering.

Collaborative Filtering algorithm considers “User Behavior” for recommending items. They exploit behavior of other users and items in terms of transaction history, ratings, selection and purchase information. Other users behaviour and preferences over the items are used to recommend items to the new users. In this case, features of the items are not known.

**Q24.** **What do you understand by Type I vs Type II error ?**

**A24.** Type I error is committed when the null hypothesis is true and we reject it, also known as a ‘False Positive’. Type II error is committed when the null hypothesis is false and we accept it, also known as ‘False Negative’.

In the context of confusion matrix, we can say Type I error occurs when we classify a value as positive (1) when it is actually negative (0). Type II error occurs when we classify a value as negative (0) when it is actually positive(1).

**Q25.** **You have been asked to evaluate a regression model based on R², adjusted R² and tolerance. What will be your criteria?**

**A25.** Tolerance (1 / VIF) is used as an indicator of multicollinearity. It is an indicator of percent of variance in a predictor which cannot be accounted by other predictors. Large values of tolerance is desirable.

We will consider adjusted R² as opposed to R² to evaluate model fit because R² increases irrespective of improvement in prediction accuracy as we add more variables. But, adjusted R² would only increase if an additional variable improves the accuracy of model, otherwise stays same. It is difficult to commit a general threshold value for adjusted R² because it varies between data sets. For example: a gene mutation data set might result in lower adjusted R² and still provide fairly good predictions, as compared to a stock market data where lower adjusted R² implies that model is not good.

**Q26.** **Considering the long list of machine learning algorithm, given a data set, how do you decide which one to use?**

**A26.** You should say, the choice of machine learning algorithm solely depends of the type of data. If you are given a data set which is exhibits linearity, then linear regression would be the best algorithm to use. If you given to work on images, audios, then neural network would help you to build a robust model.

If the data comprises of non linear interactions, then a boosting or bagging algorithm should be the choice. If the business requirement is to build a model which can be deployed, then we’ll use regression or a decision tree model (easy to interpret and explain) instead of black box algorithms like SVM, GBM etc.

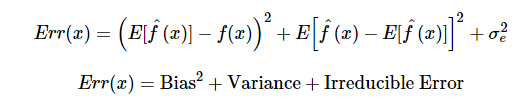
In short, there is no one master algorithm for all situations. We must be scrupulous enough to understand which algorithm to use.

**Q27.** **When does regularization becomes necessary in Machine Learning?**

**A27.** Regularization becomes necessary when the model begins to ovefit / underfit. This technique introduces a cost term for bringing in more features with the objective function. Hence, it tries to push the coefficients for many variables to zero and hence reduce cost term. This helps to reduce model complexity so that the model can become better at predicting (generalizing).

**Q28.** **What do you understand by Bias Variance trade off?**

**A28.**  The error emerging from any model can be broken down into three components mathematically. Following are these component :

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/07/error-of-a-model.png)

**Bias error** is useful to quantify how much on an average are the predicted values different from the actual value. A high bias error means we have a under-performing model which keeps on missing important trends. **Variance** on the other side quantifies how are the prediction made on same observation different from each other. A high variance model will over-fit on your training population and perform badly on any observation beyond training.

**Q29.** **OLS is to linear regression. Maximum likelihood is to logistic regression. Explain the statement.**

**A29.** OLS and Maximum likelihood are the methods used by the respective regression methods to approximate the unknown parameter (coefficient) value. In simple words,

Ordinary least square(OLS) is a method used in linear regression which approximates the parameters resulting in minimum distance between actual and predicted values. Maximum Likelihood helps in choosing the the values of parameters which maximizes the likelihood that the parameters are most likely to produce observed data.

**Q30. Python or R – Which one would you prefer for text analytics?**

A30. The best possible answer for this would be Python because it has Pandas library that provides easy to use data structures and high performance data analysis tools.

Q31. Explain the various benefits of R language?

A31. The R programming language includes a set of software suite that is used for graphical representation, statistical computing, data manipulation and calculation.  
Some of the highlights of R programming environment include the following:

* An extensive collection of tools for data analysis
* Operators for performing calculations on matrix and array
* Data analysis technique for graphical representation
* A highly developed yet simple and effective programming language
* It extensively supports machine learning applications
* It acts as a connecting link between various software, tools and datasets
* Create high quality reproducible analysis that is flexible and powerful
* Provides a robust package ecosystem for diverse needs
* It is useful when you have to solve a data-oriented problem

**Q32.** **Is rotation necessary in PCA? If yes, Why? What will happen if you don’t rotate the components?**

**A32.** Yes, rotation (orthogonal) is necessary because it maximizes the difference between variance captured by the component. This makes the components easier to interpret. Not to forget, that’s the motive of doing PCA where, we aim to select fewer components (than features) which can explain the maximum variance in the data set. By doing rotation, the relative location of the components doesn’t change, it only changes the actual coordinates of the points.

If we don’t rotate the components, the effect of PCA will diminish and we’ll have to select more number of components to explain variance in the data set.

Q33. How can you prove that one improvement you've brought to an algorithm is really an improvement over not doing anything?

A33. Often it is observed that in the pursuit of rapid innovation (aka "quick fame"), the principles of scientific methodology are violated leading to misleading innovations, i.e. appealing insights that are confirmed without rigorous validation. One such scenario is the case that given the task of improving an algorithm to yield better results, you might come with several ideas with potential for improvement.   
  
An obvious human urge is to announce these ideas ASAP and ask for their implementation. When asked for supporting data, often limited results are shared, which are very likely to be impacted by selection bias (known or unknown) or a misleading global minima (due to lack of appropriate variety in test data).   
  
Data scientists do not let their human emotions overrun their logical reasoning. While the exact approach to prove that one improvement you've brought to an algorithm is really an improvement over not doing anything would depend on the actual case at hand, there are a few common guidelines:

* Ensure that there is no selection bias in test data used for performance comparison
* Ensure that the test data has sufficient variety in order to be symbolic of real-life data (helps avoid overfitting).
* Ensure that "controlled experiment" principles are followed i.e. while comparing performance, the test environment (hardware, etc.) must be exactly the same while running original algorithm and new algorithm
* Ensure that the results are repeatable with near similar results
* Examine whether the results reflect local maxima/minima or global maxima/minima

   
One common way to achieve the above guidelines is through A/B testing, where both the versions of algorithm are kept running on similar environment for a considerably long time and real-life input data is randomly split between the two. This approach is particularly common in Web Analytics.

### Q34. Explain what resampling methods are and why they are useful. Also explain their limitations.

A34. Classical statistical parametric tests compare observed statistics to theoretical sampling distributions. Resampling a data-driven, not theory-driven methodology which is based upon repeated sampling within the same sample.   
  
Resampling refers to methods for doing one of these

* Estimating the precision of sample statistics (medians, variances, percentiles) by using subsets of available data (jackknifing) or drawing randomly with replacement from a set of data points (bootstrapping)
* Exchanging labels on data points when performing significance tests (permutation tests, also called exact tests, randomization tests, or re-randomization tests)
* Validating models by using random subsets (bootstrapping, cross validation)

### Q35. Is it better to have too many false positives, or too many false negatives? Explain.

A35. It depends on the question as well as on the domain for which we are trying to solve the question.   
  
In medical testing, false negatives may provide a falsely reassuring message to patients and physicians that disease is absent, when it is actually present. This sometimes leads to inappropriate or inadequate treatment of both the patient and their disease. So, it is desired to have too many false positive.   
  
For spam filtering, a false positive occurs when spam filtering or spam blocking techniques wrongly classify a legitimate email message as spam and, as a result, interferes with its delivery. While most anti-spam tactics can block or filter a high percentage of unwanted emails, doing so without creating significant false-positive results is a much more demanding task. So, we prefer too many false negatives over many false positives.

### Q36. What is selection bias, why is it important and how can you avoid it?

A36. Selection bias, in general, is a problematic situation in which error is introduced due to a non-random population sample. For example, if a given sample of 100 test cases was made up of a 60/20/15/5 split of 4 classes which actually occurred in relatively equal numbers in the population, then a given model may make the false assumption that probability could be the determining predictive factor. Avoiding non-random samples is the best way to deal with bias; however, when this is impractical, techniques such as [resampling](https://en.wikipedia.org/wiki/Resampling_(statistics)), [boosting](https://en.wikipedia.org/wiki/Boosting_(machine_learning)), and weighting are strategies which can be introduced to help deal with the situation.

**Q37. Differentiate between univariate, bivariate and multivariate analysis.**

A37. These are descriptive statistical analysis techniques which can be differentiated based on the number of variables involved at a given point of time. For example, the pie charts of sales based on territory involve only one variable and can be referred to as univariate analysis.

If the analysis attempts to understand the difference between 2 variables at time as in a scatterplot, then it is referred to as bivariate analysis. For example, analysing the volume of sale and a spending can be considered as an example of bivariate analysis.

Analysis that deals with the study of more than two variables to understand the effect of variables on the responses is referred to as multivariate analysis.

Q38. What is the difference between Cluster and Systematic Sampling?

A38. Cluster sampling is a technique used when it becomes difficult to study the target population spread across a wide area and simple random sampling cannot be applied. Cluster Sample is a probability sample where each sampling unit is a collection, or cluster of elements. Systematic sampling is a statistical technique where elements are selected from an ordered sampling frame. In systematic sampling, the list is progressed in a circular manner so once you reach the end of the list,it is progressed from the top again. The best example for systematic sampling is equal probability method.

Q39. Can you cite some examples where both false positive and false negatives are equally important?

A39. In the banking industry giving loans is the primary source of making money but at the same time if your repayment rate is not good you will not make any profit, rather you will risk huge losses.

Banks don’t want to lose good customers and at the same point of time they don’t want to acquire bad customers. In this scenario both the false positives and false negatives become very important to measure.

Q40. How machine learning is deployed in real world scenarios?

A40. Here are some of the scenarios in which machine learning finds applications in real world:

* Ecommerce: Understanding the customer churn, deploying targeted advertising, remarketing
* Search engine: Ranking pages depending on the personal preferences of the searcher
* Finance: Evaluating investment opportunities & risks, detecting fraudulent transactions
* Medicare: Designing drugs depending on the patient’s history and needs
* Robotics: Machine learning for handling situations that are out of the ordinary
* Social media: Understanding relationships and recommending connections
* Extraction of information: framing questions for getting answers from databases over the web

Sources:

[http://www.kdnuggets.com](http://www.kdnuggets.com/2016/02/21-data-science-interview-questions-answers.html)

[https://www.analyticsvidhya.com](https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/)

<http://analyticsindiamag.com>