42578 Advanced Business Analytics

Performance Metrics and Business Thinking

Stanislav Borysov [stabo@dtu.dk]

Adoption of Advanced Analytics Methods

Gartner’s predictions (2019)

Andrew White

Research VP

8 years at Gartner

22 years IT industry

Andrew White is a Distinguished Analyst and VP. His roles include Chief of Research and Content Lead for Data and Analytics. His main research focus is data and analytics strategy, platforms, and governance. Read Full Bio

https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

Gartner’s predictions (2019)

- By 2022, 90% of corporate strategies will explicitly mention **information** as a **critical enterprise asset** and **analytics** as an **essential competency**.

- By 2023, **data literacy** will become an **explicit and necessary driver of business value**, demonstrated by its formal inclusion in over 80% of data and analytics strategies and change management programs.

- By 2022, **more than half** of major new **business systems** will incorporate **continuous intelligence** that uses real-time context data to improve decisions.

https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

Gartner’s predictions (2019)

- By 2022, 30% of **consumers** in mature markets **will rely on artificial intelligence** (AI) to decide what they eat, what they wear or where they live.
- By 2022, 30% of organizations will use **explainable AI models** to build trust with business stakeholders, up from almost no usage today.
- By 2023, over 75% of large organizations will hire **AI behavior forensic, privacy** and **customer trust specialists** to reduce brand and reputation risk.

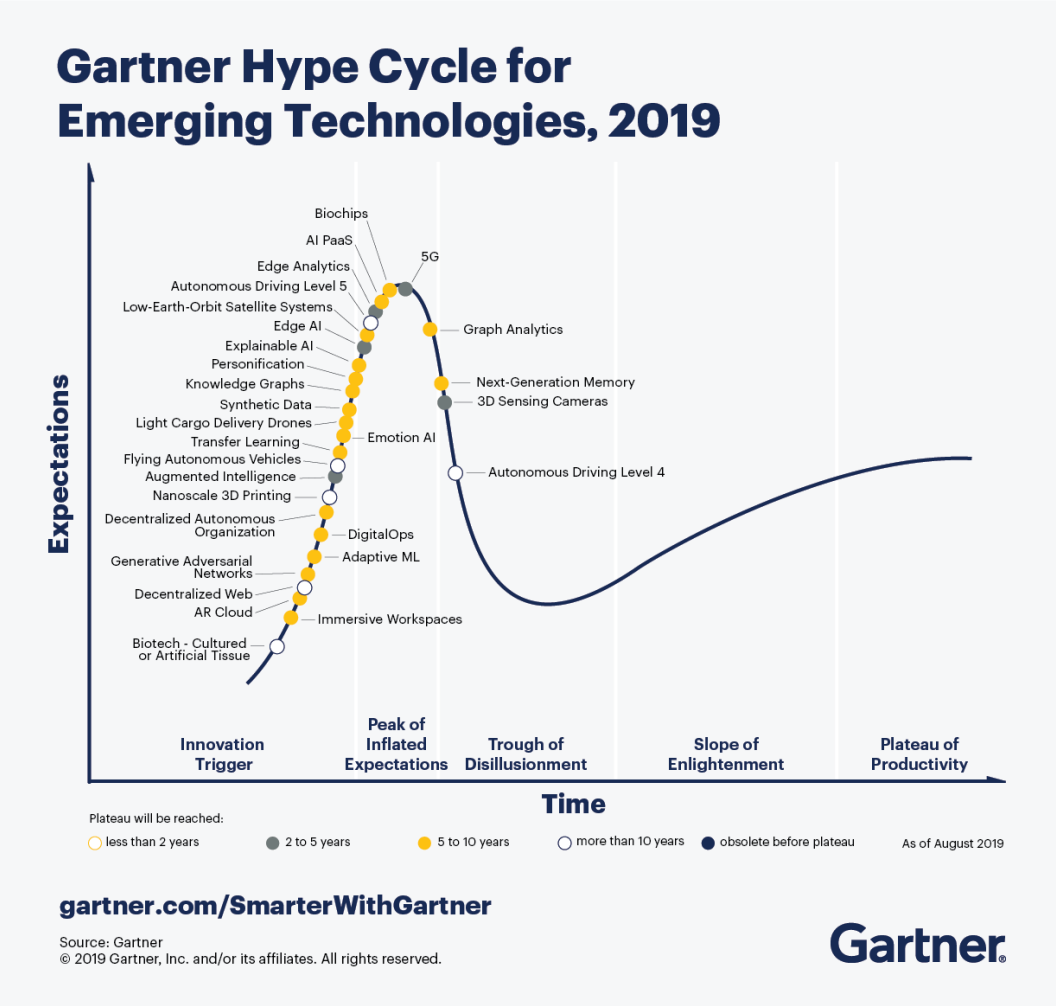
https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

Gartner’s predictions (2019)

- Through 2020, **80% of AI projects will remain alchemy**, run by wizards whose talents will not scale in the organization.
- Through 2022, **only 20% of analytic insights will deliver business outcomes**.
- Through 2022, **only 15%** of use cases leveraging **AI techniques** (such as ML and DNNs) and involving edge and IoT environments **will be successful**.

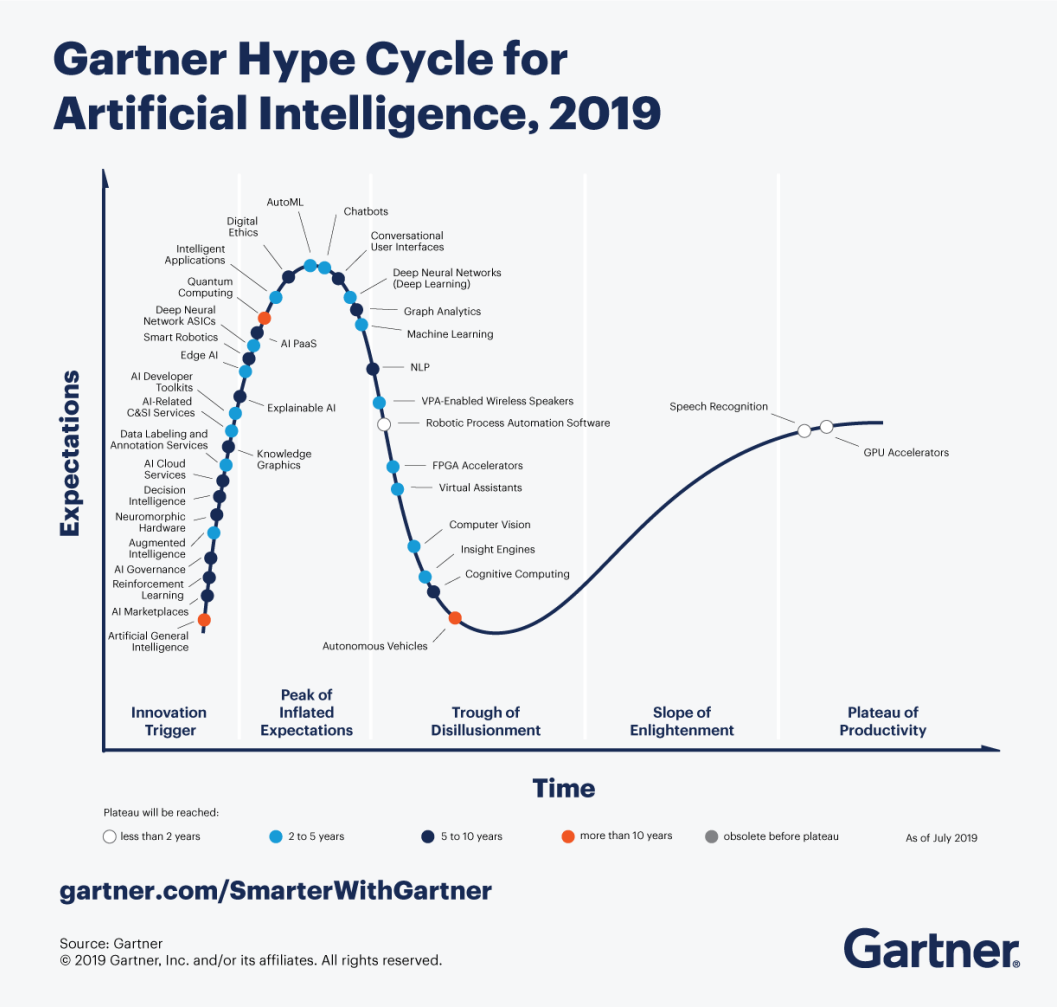
https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

Emerging Technologies Hype



https://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/

AI Hype



https://www.gartner.com/smarterwithgartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019/

Barriers to adoption of advanced BA methods

https://www.gartner.com/smarterwithgartner/3-barriers-to-ai-adoption/

Barriers to adoption of advanced BA methods

80% of companies reported stalled AI and ML projects. The following five factors are causing these slowdowns, each with its own set of challenges and opportunities:

Making data actionable for data science	Shortage of data science talent	Time-to-value must accelerate	Business users need transparency	Improving the operationalization process
• Data quality and data labeling are crucial problems slowing their AI and ML adoption. • Data silos are especially troublesome for data science. • Data science teams struggle to access and collect relevant data from different sources.	• Lack of over 150,000 people in the U.S. with data science skills. • The rapid adoption of ML and AI and the shortage of labor exacerbate the talent problem. • Finding a multi-talented unicorn is impossible. • New tech to automate and accelerate the data science process also promise to reduce talent constraints.	• Data science projects are iterative in nature due to the uncertainty of data and require a deep understanding of underlying business problems. • The process requires a rigorous trial and error approach to find the right answers, often involving multiple exchanges between business and data science teams prolonging projects.	• Frequent disconnect between ML models and the expectations of business users. • The difficulty in explaining how ML and AI models work, and how they generate results leads to a lack of trust by line of business users who don't have enough transparency to trust the process.	• Migration of data science models to production environments is fraught with impediments and challenges • Models that worked well in development don't scale and often don't work in production systems. • The result is the slow and tedious rework and "fine-tuning" of models.

@storybydata https://www.informationweek.com/big-data/five-factors-shaping-data-science/a/d-id/1335387

https://twitter.com/StorybyData/status/1161289266665799682/photo/1

Measuring Business Performance

What is a performance indicator?

- A performance indicator is a quantifiable measurement or data point used to gauge performance relative to some goal. As an example, some online retailers may have a goal to increase site traffic 50% in the next year.
- Relative to this goal, a performance indicator might be the number of unique visitors the site receives daily or which traffic sources send visitors (paid advertising, search engine optimisation, brand or display advertising, a YouTube video, etc.)

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

What is a key performance indicator?

- For most goals there could be many performance indicators — often too many — so often people narrow it down to just two or three impactful data points known as key performance indicators (KPIs). KPIs are those measurements that most accurately and succinctly show whether or not a business in progressing toward its goal.
- KPIs are like milestones on the road to business success. KPIs should be chosen and monitored depending on your unique business goals.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Why are key performance indicators important?

- KPIs are important just like strategy and goal setting are important. Without KPIs, it's difficult to gauge progress over time. You'd be making decisions based on gut instinct, personal preference or belief, or other unfounded hypotheses. KPIs tell you more information about your business and your customers, so you can make informed and strategic decisions.
- But KPIs aren't important on their own. The real value lies in the actionable insights you take away from analyzing the data. You'll be able to more accurately devise strategies to drive more online sales, as well as understand where there may problems in your business.
- Plus, the data related to KPIs can be distributed to the larger team. This can be used to educate your employees and come together for critical problem-solving.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Example: How to increase revenue?

- How can different KPIs can help you **generate more revenue** from your online shop? Here are a few ways:
  - **Increase conversion rate:** By increasing conversion rate, you are encouraging more visitors to make a purchase from your online shop.
  - **Increase the average order value:** You can entice your customers to spend more on your store when they make a purchase.
  - **Reduce the cost per acquisition:** For a sustainable and profitable business, you need to reduce the advertising and other costs you spend for acquiring a customer.

https://www.monsterinsights.com/crucial-ecommerce-kpis-to-track-in-google-analytics/

Types of KPIs in ecommerce

- There are many types of key performance indicators. They may be qualitative, quantitative, predictive of the future, or revealing of the past. KPIs also touch on various business operations. When it comes to ecommerce, KPIs generally fall into one of the following five categories:
    1. Sales
    2. Marketing
    3. Customer service
    4. Manufacturing
    5. Project management

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

1. What are KPIs for sales?

- Sales key performance indicators are measures that tell you how your business is doing in terms of conversions and revenue. You can look at sales KPIs related to a specific channel, time period, team, employee, etc. to inform business decisions.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Examples of KPIs for sales

- **Sales:** Ecommerce retailers can monitor total sales by the hour, day, week, month, quarter, or year.
- **Average order size:** Sometimes called average market basket, the average order size tells you how much a customer typically spends on a single order.
- **Gross profit:** Calculate this KPI by subtracting the total cost of goods sold from total sales.
- **Average margin:** Average margin, or average profit margin, is a percentage that represents your profit margin over a period of time.
- **Number of transactions:** This is the total number of transactions. Use this KPI in conjunction with average order size or total number of site visitors for deeper insights.
- **Conversion rate:** The conversion rate, also a percentage, is the rate at which users on your ecommerce site are converting (or buying). This is calculated by dividing the total number of visitors (to a site, page, category, or selection of pages) by the total number of conversions.
- **Shopping cart abandonment rate:** The shopping cart abandonment rate tells you how many users are adding products to their shopping cart but not checking out. The lower this number, the better. If your cart abandonment rate is high, there may be too much friction in the checkout process.
- **New customer orders vs. returning customer orders:** This metric shows a comparison between new and repeat customers. Many business owners focus only on customer acquisition, but customer retention can also drive loyalty, word of mouth marketing, and higher order values.
- **Cost of goods sold (COGS):** COGS tells you how much you're spending to sell a product. This includes manufacturing, employee wages, and overhead costs.
- **Total available market relative to a retailer's share of market:** Tracking this KPI will tell you how much your business is growing compared to others within your industry.
- **Product affinity:** This KPI tells you which products are purchased together. This can and should inform cross-promotion strategies.
- **Product relationship:** This is which products are viewed consecutively. Again, use this KPI to formulate effective cross-selling tactics.
- **Inventory levels:** This KPI could tell you how much stock is on hand, how long product is sitting, how quickly product is selling, etc.
- **Competitive pricing:** It's important to gauge your success and growth against yourself and against your competitors. Monitor your competitors' pricing strategies and compare them to your own.
- **Customer lifetime value (CLV):** The CLV tells you how much a customer is worth to your business over the course of their relationship with your brand. You want to increase this number over time through strengthening relationships and focusing on customer loyalty.
- **Revenue per visitor (RPV):** RPV gives you an average of how much a person spends during a single visit to your site. If this KPI is low, you can view website analytics to see how you can drive more online sales.
- **Churn rate:** For an online retailer, the churn rate tells you how quickly customers are leaving your brand or canceling/failing to renew a subscription with your brand.
- **Customer acquisition cost (CAC):** CAC tells you how much your company spends on acquiring a new customer. This is measured by looking at your marketing spend and how it breaks down per individual customer.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

2. What are KPIs for marketing?

- Key performance indicators for marketing tell you how well you're doing in relation to your marketing and advertising goals. These also impact your sales KPIs. Marketers use KPIs to understand which products are selling, who's buying them, how they're buying them, and why they're buying them. This can help you market more strategically in the future and inform product development.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Examples of KPIs for marketing

- **Site traffic:** Site traffic refers to the total number of visits to your ecommerce site. More site traffic means more users are hitting your store.
- **New visitors vs. returning visitors:** New site visitors are first-time visitors to your site. Returning visitors, on the other hand, have been to your site before. While looking at this metric alone won't reveal much, it can help ecommerce retailers gauge success of digital marketing campaigns. If you're running a retargeted ad, for example, returning visitors should be higher.
- **Time on site:** This KPI tells you how much time visitors are spending on your website. Generally, more time spent means they've had deeper engagements with your brand. Usually, you'll want to see more time spent on blog content and landing pages and less time spent through the checkout process.
- **Bounce rate:** The bounce rate tells you how many users exit your site after viewing only one page. If this number is high, you'll want to investigate why visitors are leaving your site instead of exploring.
- **Pageviews per visit:** Pageviews per visit refers to the average number of pages a user will view on your site during each visit. Again, more pages usually means more engagement. However, if it's taking users too many clicks to find the products they're looking for, you want to revisit your site design.
- **Average session duration:** The average amount of time a person spends on your site during a single visit is called the average session duration.
- **Traffic source:** The traffic source KPI tells you where visitors are coming from or how they found your site. This will provide information about which channels are driving the most traffic, such as: organic search, paid ads, or social media.
- **Mobile site traffic:** Monitor the total number of users who use mobile devices to access your store and make sure your site is optimized for mobile.
- **Newsletter subscribers:** The number of newsletter subscribers refers to how many users have opted into your email marketing list. If you have more subscribers, you can reach more consumers. However, you'll also want to look at related data, such as the demographics of your newsletter subscribers, to make sure you're reaching your target audience.
- **Subscriber growth rate:** This tells you how quickly your subscriber list is growing. Pairing this KPI with the total number of subscribers will give you good insight into this channel.
- **Email open rate:** This KPI tells you the percentage of subscribers that open your email. If you have a low email open rate, you could test new subject lines, or try cleaning your list for inactive or irrelevant subscribers.
- **Email click-through rate (CTR):** While the open rate tells you the percentage of subscribers who open the email, the click-through rate tells you the percentage of those who actually clicked on a link after opening. This is arguably more important than the open rate because without clicks, you won't drive any traffic to your site.
- **Unsubscribes:** You can look at both the total number and the rate of unsubscriptions for your email list.
- **Social followers and fans:** Whether you're on Facebook, Instagram, Twitter, Pinterest, or Snapchat (or a combination of a few), the number of followers or fans you have is a useful KPI to gauge customer loyalty and brand awareness. Many of those social media networks also have tools that ecommerce businesses can use to learn more about their social followers.
- **Social media engagement:** Social media engagement tells you how actively your followers and fans are interacting with your brand on social media.
- **Average CTR:** The average click-through rate tells you the percentage of users on a page (or asset) who click on a link.
- **Average position:** The average position KPI tells you about your site's search engine optimization (SEO) and paid search performance. This demonstrates where you are on search engine results pages. Most online retailers have the goal of being number one for their targeted keywords.
- **Pay-per-click (PPC) traffic volume:** If you're running PPC campaigns, this tells you how much traffic you're successfully driving to your site.
- **Number and quality of product reviews:** Product reviews are great for a number of reasons: They provide social proof, they can help with SEO, and they give you valuable feedback for your business. The quantity and content of product reviews are important KPIs to track for your ecommerce business.
- **Banner or display advertising CTRs:** The CTRs for your banner and display ads will tell you the percentage of viewers who have clicked on the ad. This KPI will give you insight into your copy, imagery, and offer performance.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

3. What are KPIs for customer service?

- Customer service KPIs tell you how effective your customer service is and if you’re meeting expectations.You might be wondering: what should the KPIs be in our call center, for our email support team, for our social media support team, etc. Measuring and tracking these KPIs will help you ensure you’re providing a positive customer experience.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Examples of KPIs for customer service

- **Customer satisfaction (CSAT) score:** The CSAT KPI is typically measured by customer responses to a very common survey question: “How satisfied were you with your experience?” This is usually answered with a numbered scale.
- **Net promoter score (NPS):** Your NPS KPI provides insight into your customer relationships and loyalty by telling you how likely customers are to recommend your brand to someone in their network.
- **Hit rate:** Calculate your hit rate by taking the total number of sales of a single product and dividing it by the number of customers who have contacted your customer service team about said product.
- **Customer service email count:** This is the number of emails your customer support team receives.
- **Customer service phone call count:** Rather than email, this is how frequently your customer support team is reached via phone.
- **Customer service chat count:** If you have live chat on your ecommerce site, you may have a customer service chat count.
- **First response time:** First response time is the average amount of time it takes a customer to receive the first response to their query. Aim low!
- **Average resolution time:** This is the amount of time it takes for a customer support issue to be resolved, starting from the point at which the customer reached out about the problem.
- **Active issues:** The total number of active issues tells you how many queries are currently in progress.
- **Backlogs:** Backlogs are when issues are getting backed up in your system. This could be caused by a number of factors.
- **Concern classification:** Beyond the total number of customer support interactions, look at quantitative data around trends to see if you can be proactive and reduce customer support queries. You’ll classify the customer concerns which will help identify trends and your progress in solving issues.
- **Service escalation rate:** The service escalation rate KPI tells you how many times a customer has asked a customer service representative to redirect them to a supervisor or other senior employee. You want to keep this number low.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

4. What are KPIs for manufacturing?

- Key performance indicators for manufacturing are, predictably, related to your supply chain and production processes. These may tell you where efficiencies and inefficiencies are, as well as help you understand productivity and expenses.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Examples of KPIs for manufacturing

- **Cycle time:** The cycle time manufacturing KPI tells you how long it takes to manufacture a single product from start to finish. Monitoring this KPI will give you insight into production efficiency.
- **Overall equipment effectiveness (OEE):** The OEE KPI provides ecommerce businesses with insight into how well manufacturing equipment is performing.
- **Overall labor effectiveness (OLE):** Just as you'll want insight into your equipment, the OLE KPI will tell you how productive the staff operating the machines are.
- **Yield:** Yield is a straightforward manufacturing KPI. It is the number of products you have manufactured. Consider analyzing the yield variance KPI in manufacturing, too, as that will tell you how much you deviate from your average.
- **First time yield (FTY) and first time through (FTT):** FTY, also referred to as first pass yield, is a quality-based KPI. It tells you how wasteful your production processes are. To calculate FTY, divide the number of successfully manufactured units by the total number of units that started the process.
- **Number of non-compliance events or incidents:** In manufacturing, there are several sets of regulations, licenses, and policies businesses must comply with. These are typically related to safety, working conditions, and quality. You'll want to reduce this number to ensure you're operating within the mandated guidelines.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

5. What are KPIs for project management?

- Key performance indicators for project management give you insight into how well your teams are performing and completing specific tasks. Each project or initiative within your ecommerce business has different goals, and must be managed with different processes and workflows. Project management KPIs tell you how well each team is working to achieve their respective goals and how well their processes are working to help them achieve those goals.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

KPIs for project management

- **Hours worked:** The total hours worked tells you how much time a team put into a project. Project managers should also assess the variance in estimated vs. actual hours worked to better predict and resource future projects.
- **Budget:** The budget indicates how much money you have allocated for the specific project. Project managers and ecommerce business owners will want to make sure that the budget is realistic; if you're repeatedly over budget, some adjustments to your project planning need to be made.
- **Return on investment (ROI):** The ROI KPI for project management tells you how much your efforts earned your business. The higher this number, the better. The ROI accounts for all of your expenses and earnings related to a project.
- **Cost variance:** Just as it's helpful to compare real vs. predicted timing and hours, you should examine the total cost against the predicted cost. This will help you understand where you need to reel it in and where you may want to invest more.
- **Cost performance index (CPI):** The CPI for project management, like ROI, tells you how much your resource investment is worth. The CPI is calculated by dividing the earned value by the actual costs. If you come in under one, there's room for improvement.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

How do I create a KPI?

- Selecting your KPIs begins with clearly stating your goals and understanding which areas of business impact those goals. Of course, KPIs for ecommerce can and should differ for each of your goals, whether they're related to boosting sales, streamlining marketing, or improving customer service.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Key performance indicator templates

- **GOAL 1: Boost sales 10% in the next quarter.**
- **KPI examples:**
  - Daily sales.
  - Conversion rate.
  - Site traffic.

- **GOAL 2: Increase conversion rate 2% in the next year.**
- **KPI examples:**
  - Conversion rate.
  - Shopping cart abandonment rate.
  - Competitive pricing.

- **GOAL 3: Grow site traffic 20% in the next year.**
- **KPI examples:**
  - Site traffic.
  - Traffic sources.
  - Promotional click-through rates.
  - Social shares.
  - Bounce rates.

- **GOAL 4: Reduce customer service calls by half in the next 6 months.**
- **KPI examples:**
  - Service call classification.
  - Pages visited immediately before call.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

Actions to improve KPIs

- Once you have set goals and selected KPIs, monitoring those indicators should become an everyday exercise.
- Most importantly: Performance should inform business decisions and you should use KPIs to drive actions.

https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce

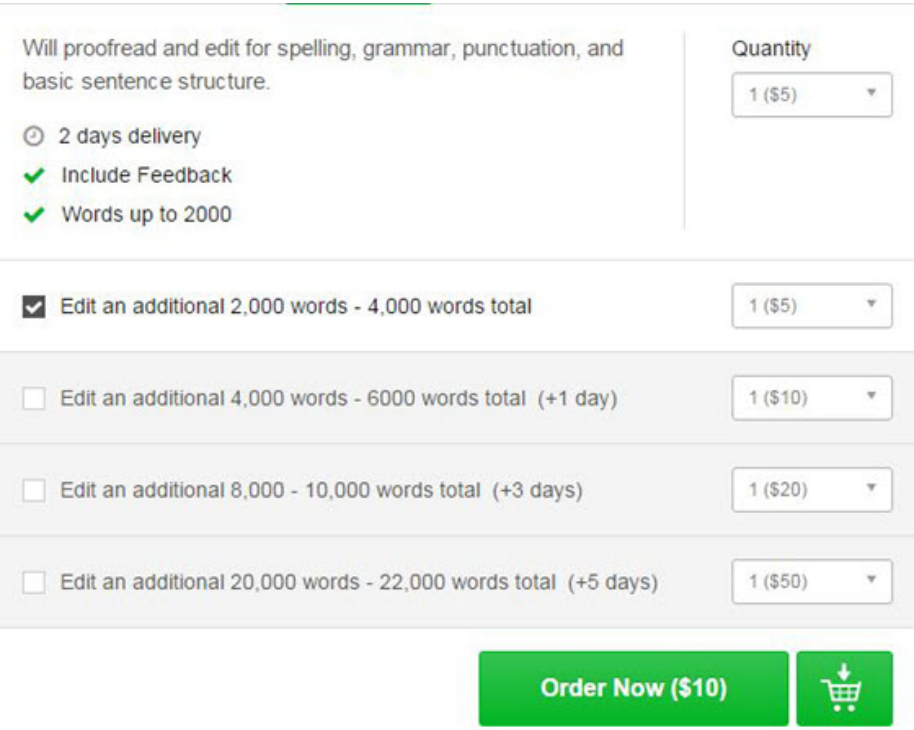
Example: Average Order Value

- One of the best ways to increase the revenue you generate from your store is by encouraging your customers to spend more when they make a purchase.
- To entice customers to spend more on your store, you need to focus on tracking and improving your Average Order Value (AOV).
- The higher the AOV, the more income your store generates on each order. The basic calculation is the total sum of revenue generated divided by the total number of orders.

https://www.monsterinsights.com/crucial-ecommerce-kpis-to-track-in-google-analytics/

Increasing Average Order Value

- **Upsell**
  You can persuade your customers to purchase something additional or more expensive ones.



https://www.monsterinsights.com/crucial-ecommerce-kpis-to-track-in-google-analytics/

Increasing Average Order Value

- **Sell Product bundles**
  For a discounted price, you can sell several products in a bundle.

https://www.monsterinsights.com/crucial-ecommerce-kpis-to-track-in-google-analytics/

Increasing Average Order Value

- **Offer discounts on a minimum spend threshold**
  You can convince your customers to spend more while making a purchase by setting a minimum spend threshold.

Spend $25, get $5 FREE
Spend $50, get $10 FREE
Spend $100 or more, get $20 FREE

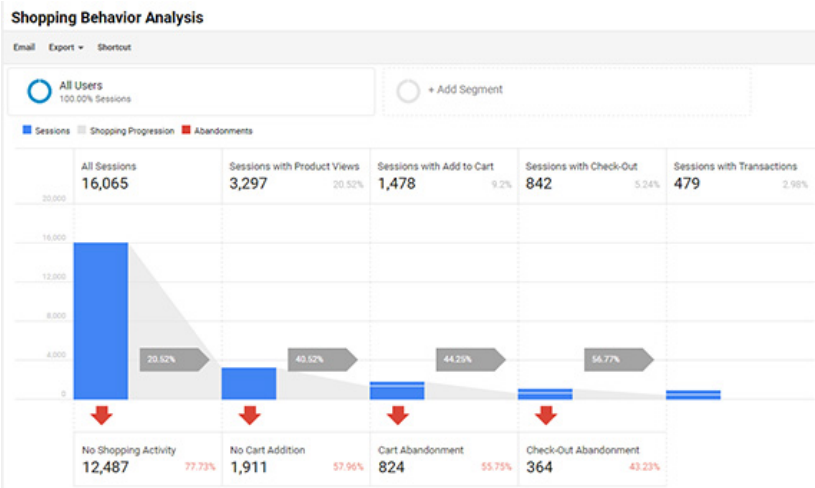
It's Fiverr's 7th birthday. But doers don't take days off, so we're spreading the LOVE with our biggest rewards yet. Doers, this deal lasts for 7 days until Feb 20, 2017. Just pick your offer.

Join the Celebration

https://www.monsterinsights.com/crucial-ecommerce-kpis-to-track-in-google-analytics/

Example: Shopping Cart Abandonment

- According to Baymard Institute, on average, 69 out of 100 orders are abandoned before being completed [https://baymard.com/lists/cart-abandonment-rate]



- **A few reasons that make customers abandon the cart are:**
  - Seeing unexpected costs on checkout.
  - Not following the promises you make.
  - No security badges or trust seals.
  - Too long registration form.

https://www.monsterinsights.com/crucial-ecommerce-kpis-to-track-in-google-analytics/

Decreasing Shopping Cart Abandonment

1. Develop new checkout system with presumably improved user experience
2. Do A/B testing
    1. Show the new design to 5% of users
    2. Measure if there is a statistically significant difference in KPIs between the old and new designs

Measuring Predictive Modelling Performance

Questions to ask

- *What do we want to optimise for?*
- *What are the costs of prediction errors?*
- *What is the ROI of a better model (development / operation costs)?*

https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226

Regression performance metrics

RMSE vs MAE

- **RMSE (Root Mean Square Error)**
  = the sample standard deviation of the differences between predicted values and observed values (residuals)

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{N}\mathrm{SE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}$$

- **MAE (Mean Absolute Error)**
  = the average of the absolute values of the residuals

$$\mathrm{MAE} = \frac{1}{N}\mathrm{AE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|$$

https://medium.com/usf-msds/choosing-the-right-metric-for-machine-learning-models-part-1-a99d7d7414e4

RMSE vs MAE: Which one to use?

- Both are measured in **the same units as the target variable,** y

- However, **MAE is less sensitive to outliers**:
  The MAE is a linear score which means that all the individual differences are weighted equally in the average. For example, the difference between 10 and 0 will be twice the difference between 5 and 0. However, same is not true for RMSE which penalises the higher difference more than MAE.

- **It is business-problem-specific:**
  Are the outliers important? Higher errors are worse?
  - Yes - use RMSE
  - No - use MAE

Other measures

- **Absolute** (the same units as y)
  RMSE, MAE, asymmetric quantile errors, …

- **Relative** (%)
  - Normalised

$$\frac{\text{RMSE}, \text{MAE}, \ldots}{\bar{y}, Q_{0.5}, \sigma_y, Q_{0.75} - Q_{0.25}, y_{\max} - y_{\min}, \ldots}, \quad \text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \quad \ldots$$

  - Coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}_i\right)^2} = 1 - \frac{\text{MSE}}{\sigma_y^2} = 1 - \left(\frac{\text{RMSE}}{\sigma_y}\right)^2$$

  - Correlation between the predicted and true values (“45-degree plot”), …

Regression: General suggestions

- **Technical context:**
  - Report quite a few performance metrics

- **Business context:**
  - Focus on 1-2 relevant performance metrics which are easy to communicate and translate into the business language
  - Try to quantify the errors, i.e., *business value = f(errors)*
    - Prediction error decreased by X → KPI increased by Y
    - Is it asymmetric? (e.g., is excess of items is better than shortage?)
    - Assumptions are needed, might be difficult to guess from the beginning (e.g., how much the company will gain if its recommender system error is lower by 1 star?)
  - In model fitting (parameters & hyperparameters), try to optimise for a more relevant score from the business perspective

Classification performance metrics

Binary classification

- A classification model f is fitted to the data to predict **score** $s^{(i)} = f\left(x^{(i)}\right)$ measuring model's belief that the sample $x^{(i)}$ belongs to the class 1 ("positive" class)
- The predicted class is based on the **decision threshold** $\tilde{s}$: $x^{(i)}$ belongs to the positive class if $s^{(i)} \geq \tilde{s}$ and to the "negative" class otherwise
  - E.g., Logistic Regression default $\tilde{s} = 0.5$ but it can be changed!
- Given the predicted scores and the threshold, a **confusion matrix** can be constructed

		Actual: Positive class	Actual: Negative class
Predicted	Positive class	True Positives (TP)	False Positives (FP) "Type I error"
Predicted	Negative class	False Negatives (FN) "Type II error"	True Negatives (TN)

Binary classification performance metrics

- Based on the scores and class thresholds, three types of performance metrics can be considered:

  1. **Ranking metrics**
     How well the model ranks the examples and separates classes
     Examples: ROC and precision recall curves, ROC AUC

  2. **Probability metrics**
     Measure the deviation of the predicted probabilities from the true ones
     Examples: cross-entropy, Brier score

  3. **Threshold metrics**
     Measure the classification prediction errors
     Examples: accuracy, precision, recall, F-measure, Cohen's kappa

- *Almost everything is implemented in sklearn*

1. Ranking metrics

- Metrics based on how well the model ranks the examples. These are important for many applications where classifiers are used to **select the top n instances** of a set of data or when good class separation is crucial.

- *Business value = f(ranking)*

- **Business cases:** Recommend items, distribute limited marketing budget among potential churns, …

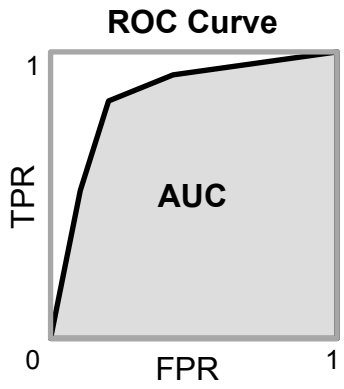
- **Most popular metrics:**
  - Receiver operating characteristic (ROC) curve
  - Precision-recall (PR) curve
  - Usually, area under a curve (AUC) is calculated to get a single number

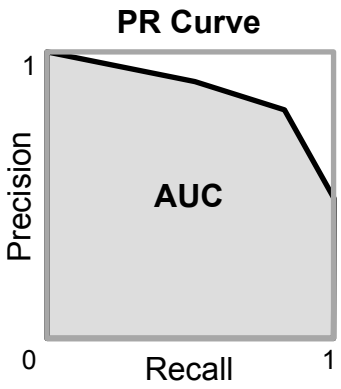
https://www.sciencedirect.com/science/article/abs/pii/S0167865508002687
https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/

	True 1	True 0
Pred 1	TP	FP
Pred 0	FN	TN

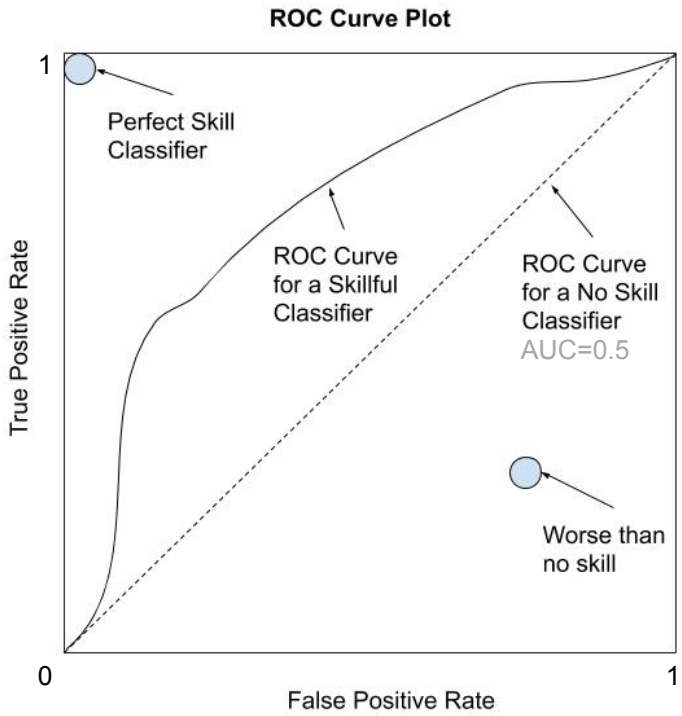
1. Ranking metrics: General approach

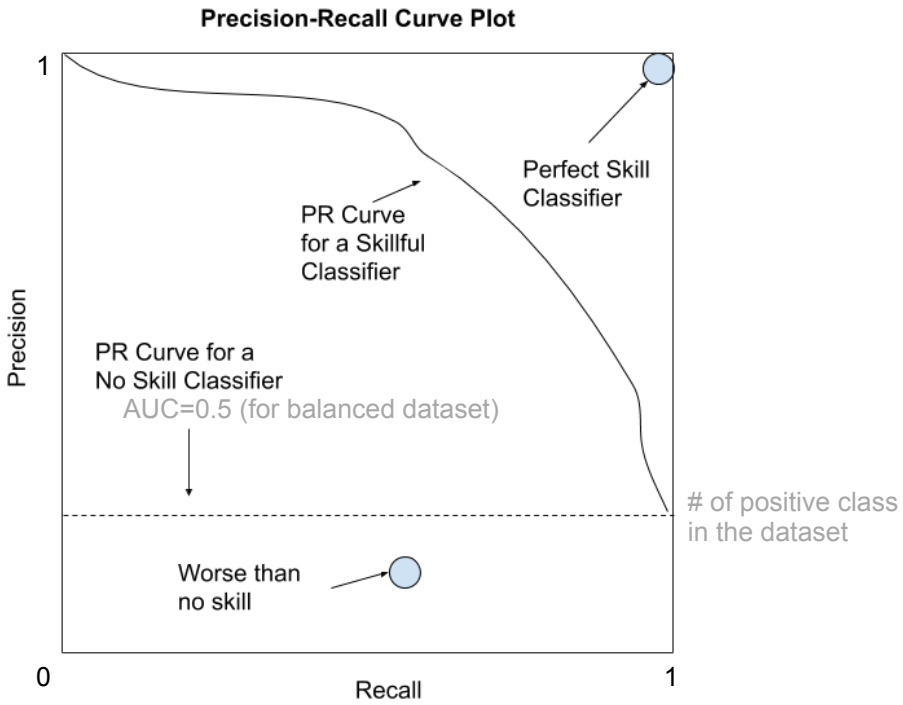
1. Predict scores using a model
2. Change decision threshold from 0 to max
    1. For each decision threshold, calculate confusion matrix
        - **For ROC, calculate**
          TPR (True Positive Rate) = Sensitivty = Recall = TP/(TP+FN)
          FPR (False Positive Rate) = 1- Specificity = FP/(TN+FP)
        - **For PR, calculate**
          Precision = TP/(TP+FP)
          Recall = TP/(TP+FN)
    2. Plot one versus another
3. To get one number, calculate area under the curve (AUC)





1. Ranking metrics: Plots





Treats two classes equally

Focused more on the positive class

https://www.sciencedirect.com/science/article/abs/pii/S0167865508002687

https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/

2. Probability metrics

- Metrics based on a probabilistic understanding of error, i.e. **measuring the deviation from the true probability**. These measures are especially useful when we want an assessment of the reliability of the classifiers, not only measuring when they fail but whether they have selected the wrong class with a high or low probability.

- *Business value = f(class probabilities)*

- **Business cases:** Take action for the potential high-probability churns only, assign dangerous treatment to a patient when we really sure about it, …

- **Most popular metrics:**
  - Binary loss/cross-entropy = log-likelihood of a categorical variable
  - Brier score

https://www.sciencedirect.com/science/article/abs/pii/S0167865508002687
https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/

2. Probability metrics

- **Binary loss/cross-entropy** = log-likelihood of a categorical variable

$$\mathrm{LL}_i = -\sum_{k=1}^{K} y_k^{(i)} \log\left(\hat{y}_k^{(i)}\right)$$

- **Brier score**

$$\mathrm{BS}_i = \sum_{k=1}^{K} \left(y_k^{(i)} - \hat{y}_k^{(i)}\right)^2$$

- **Comments:**
  - Remember to calibrate probabilities
  - In principle, both can be weighted for each class, e.g.

$$-\sum_{k=1}^{K} w_k y_k^{(i)} \log\left(\hat{y}_k^{(i)}\right)$$

https://www.sciencedirect.com/science/article/abs/pii/S0167865508002687
https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/

3. Threshold metrics

- Metrics based on a threshold and a qualitative understanding of error. These measures are used when we want a model to **minimise the number of errors**.

- *Business value = f(CM)*

- **Examples:** Almost all classification problems

- **Most popular metrics:**
  - Accuracy
  - Precision
  - Recall
  - F-score

- **Usually the most relevant type of metrics in business because it is both easy to understanding and quantify the costs associated with errors**

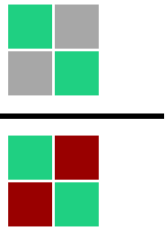
https://www.sciencedirect.com/science/article/abs/pii/S0167865508002687
https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/

	True 1	True 0
Pred 1	TP	FP
Pred 0	FN	TN

3. Threshold metrics: Accuracy

- Accuracy is the proportion of the correct predictions

- $$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} =$$



- **Comments:**
  - Both errors are equally important ~ “Making any error is costly”
  - Optimising for binary/categorical cross-entropy = optimising for accuracy
  - Usually, the default metrics
  - Works poorly for imbalanced problems (always predict the major class and get high accuracy)

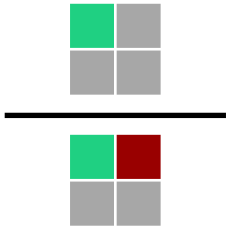
https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226

	True 1	True 0
Pred 1	TP	FP
Pred 0	FN	TN

3. Threshold metrics: Precision

- Precision is the proportion of predicted positives which are truly positive

- $$\text{Precision} = \frac{TP}{TP + FP} =$$



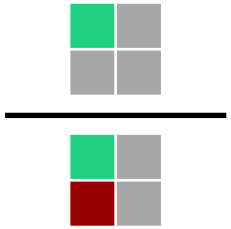
- **Comments:**
  - FP are more important ~ “Misclassifying true negatives is costly”
  - Precision is a valid choice of evaluation metric when we want to be very sure of our prediction. For example: If we are building a system to predict if we should decrease the credit limit on a particular account, we want to be very sure about our prediction or it may result in customer dissatisfaction.
  - Being very precise means our model will leave a lot of credit defaulters untouched and hence lose money.

https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226

	True 1	True 0
Pred 1	TP	FP
Pred 0	FN	TN

3. Threshold metrics: Recall

- Recall is the proportion of the true positives which are correctly predicted

- $\text{Recall} = \frac{TP}{TP + FN} =$ 

- **Comments:**
  - FN are more important ~ “Misclassifying true positives is costly”
  - Recall is a valid choice of evaluation metric when we want to capture as many positives as possible. For example: If we are building a system to predict if a person has cancer or not, we want to capture the disease even if we are not very sure.
  - Recall is 1 if we predict 1 for all examples.

https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226

	True 1	True 0
Pred 1	TP	FP
Pred 0	FN	TN

3. Threshold metrics: F-score

- F-score is a weighted harmonic average between precision and recall

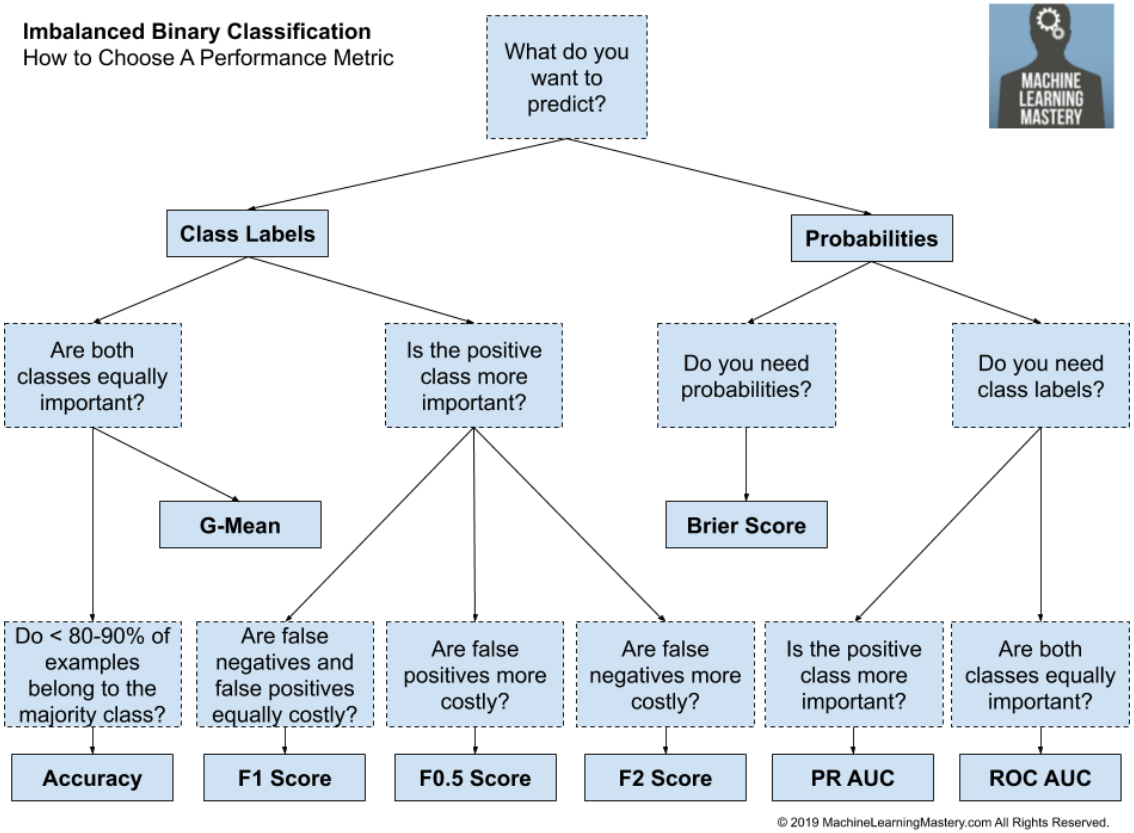
- $F_\beta = (1+\beta^2)\dfrac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}$
  - $\beta = 1 \rightarrow$ Precision and Recall are equally important
  - $\beta = 0.5 \rightarrow$ Precision is more important
  - $\beta = 2 \rightarrow$ Recall is more important

- **Comments:**
  - Simply stated the F1 score sort of maintains a balance between the precision and recall for your classifier. If your precision is low, the F1 is low and if the recall is low again your F1 score is low → **Good for imbalanced problems**
  - If you are a police inspector and you want to catch criminals, you want to be sure that the person you catch is a criminal (Precision) and you also want to capture as many criminals (Recall) as possible. The F1 score manages this tradeoff.

https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226

Imbalanced Classification

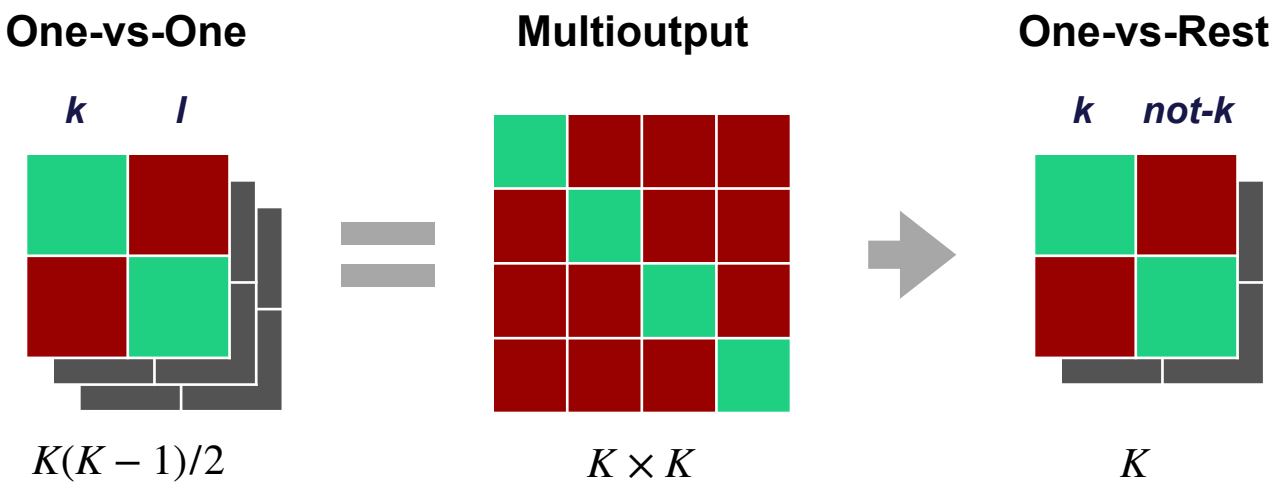
- https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/



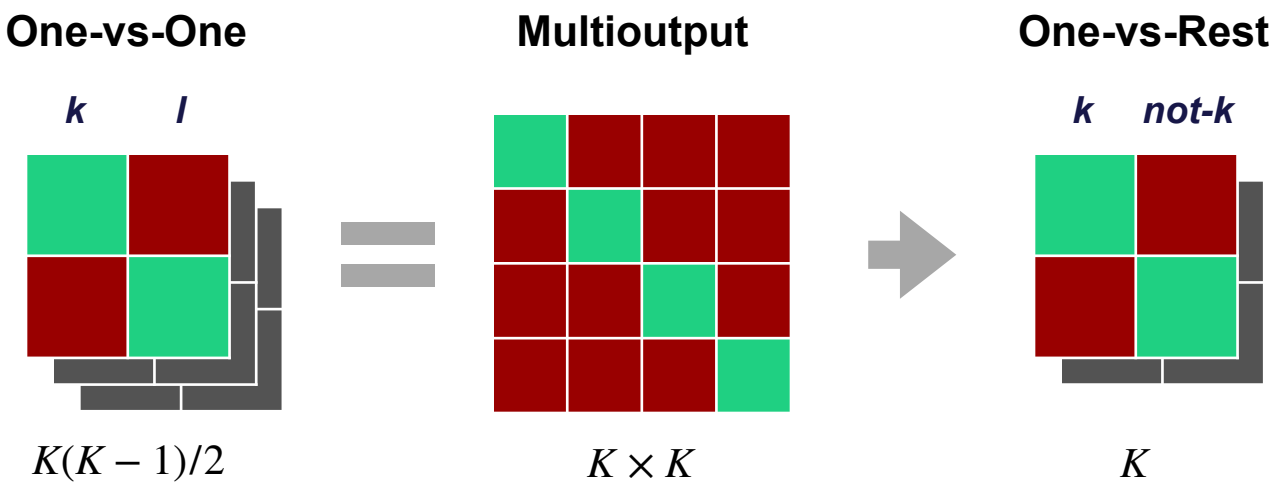
Multiclass classification (*K* classes)

- **Multiouput**: Single model f predicts **score for each class** $s_k^{(i)} = f\left(x^{(i)}\right)$ measuring model's belief that the sample $x^{(i)}$ belongs to the class k
  - The class is predicted using the maximum score $k^* = \max_k s_k^{(i)}$
    - Optionally, with weights $k^* = \max_k w_k s_k^{(i)}$
- **One-vs-rest:** Separate binary model for each class f_k (K models in total) predicts "class k" versus "not-class-k" for each class k with a decision threshold for each class $\tilde{s}_k$.
- **One-vs-one:** Separate binary model for each pair of classes f_{kl} ($K(K-1)/2$ models in total) predicts "class k" versus "class l" for each pair of classes k and l with a decision threshold for each class pair $\tilde{s}_{kl}$.

Multiclass classification (*K* classes)



Multiclass classification (*K* classes)



- How to find a single number which shows average performance of the classifier?

Multiclass classification performance metrics

- **Common way:** K one-vs-rest confusion matrices, CM_k, are considered
- Let's say, we want to calculate a single F-score. There are several options:
  - **Macro-average:** Calculate the F-score for each CM_k, then average the scores
    - Each class contributes equally to the final score independently of number of samples (good approach for imbalanced classification)
    - Alternatively, F-scores can be weighted by the number of samples
    - You can also calculate variance of the scores
  - **Macro-average*:** Calculate the macro-averaged precision and recall, then calculate the F-score
    - Possible, but rarely used on practice
  - **Micro-average:** Calculate the total confusion matrix $\mathrm{CM} = \sum_{k=1}^{K} \mathrm{CM}_k$, then calculate the F-score
    - More frequent classes will dominate ~ weighted macro-average

How to optimise a classifier for a business problem?

- **Choose a relevant metric, *M*, based on estimated cost of errors**
  - Precise cost is not known but there are some general feelings → choose something from the standard metrics, e.g., F2-score
  - Construct your own metric

- **Option 1 (most common & good enough way)**
  - Train the model optimising for cross-entropy (~ optimising for accuracy)
  - Tune HP optimising for *M*
  - Fine-tune decision boundary optimising for the final costs. You can adjust the decision boundary quite frequently even when the model is in production

- **Option 2 (advanced)**
  - Directly use *M* to train the model
    - You have to make M differentiable. There are such things as soft F-score, differentiable ROC AUC, etc.
  - The rest is the same

One last thing for regression & classification

- **… and always remember to set baselines!** For example,
  - Always predicting most frequent class
  - Predicting randomly from the data distribution (stratified)
  - Predicting mean and/or median
  - …

- https://scikit-learn.org/stable/modules/model_evaluation.html#dummy-estimators
  - `sklearn.dummy.DummyRegressor`
  - `sklearn.dummy.DummyClassifier`

- **… and translate model performance metrics into business / non-technical / human language when communicating your results to non-technical people!**

Communicating Results

R. Kim, Talking technology to non-technical people, https://slideplayer.com/slide/17587343/

Communicating results

- One of the most important skills for analysts to have is being able to clearly communicate results so different stakeholders can understand. Since data projects are collaborative across functions and advanced analytics results are often incorporated into a larger final project, the true impact of analysts' work depends on how well others can understand their insights to take further action.

https://medium.com/comet-ml/a-data-scientists-guide-to-communicating-results-c79a5ef3e9f1

Who is your audience?

- Throughout every project, you'll encounter people with varying levels of technical expertise, buy-in, and business goals. When you present to these different stakeholders, make sure you keep an eye on how your work ties into their role and decisions.
- With **non-technical stakeholders, you should avoid highly technical terms** (e.g. your hyperparameters for your TensorFlow model) and instead, try to frame the machine learning problem into the same terms in which business decisions are made, e.g., marginal cost and benefit.
- However, with your engineering or devOps team, they will need to know details such as how long the model takes to train and GPU/CPU metrics during training.

https://medium.com/comet-ml/a-data-scientists-guide-to-communicating-results-c79a5ef3e9f1

Communicating with business stakeholders

1 Put Everything in Business Terms

- Technical people tend to get excited when we get to talk about the machine learning algorithm we chose, why we chose it, and various model diagnostics. It is all too easy to geek out on how good the precision and recall of your classification model is, and if you do this you will quickly lose your audience because they likely can't relate to these concepts. Instead, **try repackaging your findings into business terms that are already understood.**

- For example, let's say you have built this great model that predicts when a piece of equipment on an oil well site is going to lead to a site shutdown. Instead of talking about how you have achieved a recall score of .95, say that you are able to detect 95% of site shutdowns caused by that particular piece of equipment. By repackaging your message into terms that are already understood, we are able to communicate the same concepts in a way that immediately resonates with a business minded audience.

https://nealanalytics.com/3-key-strategies-for-communicating-data-science-insights/

Communicating with business stakeholders

2 Place an Emphasis on Business Value

- We have already stated that we can predict 95% of shutdowns that are caused by a specific piece of equipment. So far this doesn't mean anything to the business. What if we can predict 95% of these events, but they only account for 10% of total shutdown events? What if the cost per shutdown event caused by this piece of equipment is insignificant?

- Placing an emphasis on potential business value will help put these questions to rest as well as demonstrate the need for continued use of advanced analytics. In our current scenario, we must do our due diligence to investigate things like: how many shutdown events occur per year, what proportion are caused by the equipment studied, how long an event lasts on average, and what an event costs the organisation in lost production and repair costs. Armed with this information, **we can directly address the questions of key stakeholders by emphasising the current state of the business, how we can improve this state with our solution, and what this improvement means in financial terms.**

https://nealanalytics.com/3-key-strategies-for-communicating-data-science-insights/

Communicating with business stakeholders

3 Provide Real-World Examples of Benefits

- Nearly every professional wants to know how new technology makes their lives easier. To achieve this, **provide real-world examples of the benefits of next-gen IT or advanced data sets**. While it's easy to fall into the trap of listing intimate details on every step of the process, there's a lot to be said for brevity — especially in a role such as this.

- Although it's best to demonstrate how new technology or information benefits the specific company, you can use case studies to provide supplemental predictions and forecasts. Doing so is less effective than producing models based on the company's internal processes or dataflow, but it still provides an excellent way of making your point as quickly and completely as possible.

https://insidebigdata.com/2018/03/28/data-scientists-guide-communicating-results/

Communicating with business stakeholders

4 Focus on the Result

- Try not to get bogged down with the technical details of any specific project. More importantly, don't overload your client or manager with information from the start. Instead, **begin by drawing attention to the result and work backward from there**. Instead of explaining the technical specifications or requirements of a new application or process, describe the final benefits. Not only does this capture your audience's attention immediately, but it might be enough to mitigate any other concerns and gain the necessary approval.

https://insidebigdata.com/2018/03/28/data-scientists-guide-communicating-results/

Communicating with business stakeholders

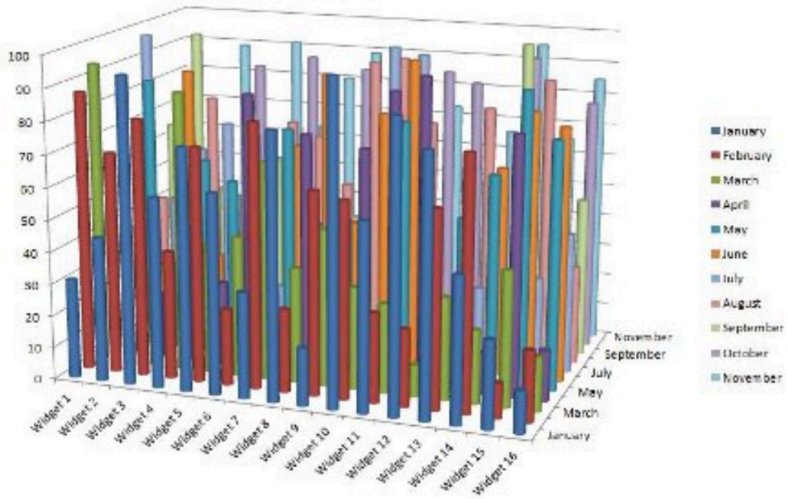
5 Use Visualisations Whenever Applicable

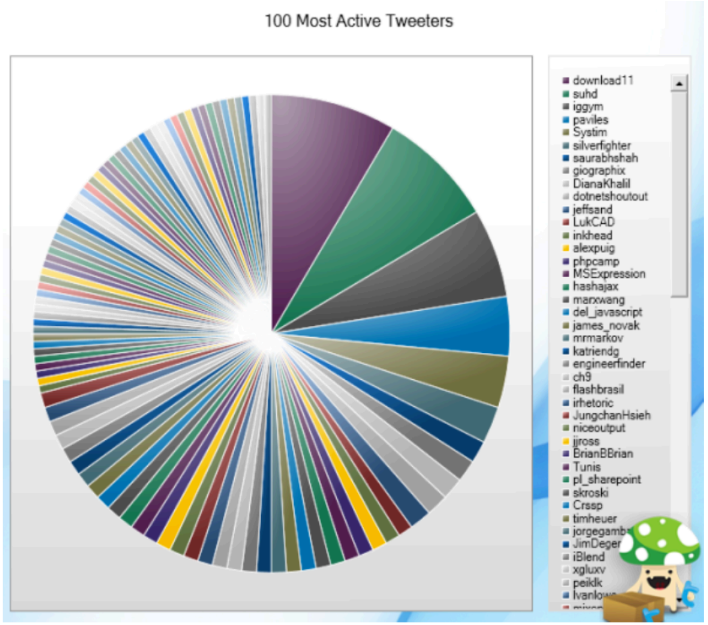
- Charts, graphs, illustrations and spreadsheets can work wonders for an otherwise boring or drab report. Today's computers, including laptops and even many smartphones, come pre-loaded with intuitive design software for this very purpose. Use these documents to your advantage whenever applicable. They make it easy to display various datasets, highlight different stats and draw attention to the most important points you're trying to make.

https://insidebigdata.com/2018/03/28/data-scientists-guide-communicating-results/

Communicating with business stakeholders

5 Use Visualisations Whenever Applicable

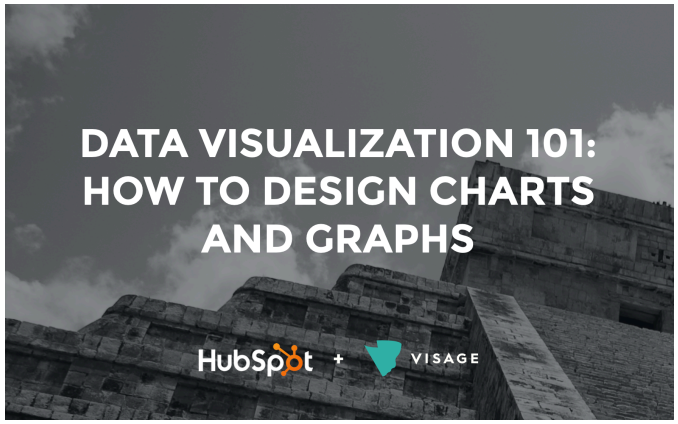




Communicating with business stakeholders

5 Use Visualisations Whenever Applicable

- **Recommended reading**
  - https://portal.yellowfinbi.com/Document.i4?DocumentId=1009768 (data_visualization_yellowfin.pdf)
  - https://cdn2.hubspot.net/hub/53/file-863940581-pdf/Data_Visualization_101_How_to_Design_Charts_and_Graphs.pdf (data_visualization_hubspot_visage.pdf)



Summary: General framework

1. Your **understanding of the business problem**
2. **How to measure business impact** — what business metrics do your model results align to?
3. **What data is/was available** — if appropriate, reference what data it would be helpful to collect
4. The **initial solution hypothesis**
5. The **solution/model** — use examples and visualizations
6. The **business impact** of the solution and **clear action items** for stakeholders

https://medium.com/comet-ml/a-data-scientists-guide-to-communicating-results-c79a5ef3e9f1

Recommended reading

- **KPIs for eCommerce**
  - https://www.shopify.com/blog/7365564-32-key-performance-indicators-kpis-for-ecommerce
  - https://www.monsterinsights.com/crucial-ecommerce-kpis-to-track-in-google-analytics/

- **Evaluation metrics for imbalanced classification**
  - https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/

- **Data visualisation tips**
  - https://portal.yellowfinbi.com/Document.i4?DocumentId=1009768 (data_visualization_yellowfin.pdf)
  - https://cdn2.hubspot.net/hub/53/file-863940581-pdf/Data_Visualization_101_How_to_Design_Charts_and_Graphs.pdf (data_visualization_hubspot_visage.pdf)
  - https://sloanreview.mit.edu/article/choose-charts-everyone-understands/