* Annexe 1

#### Analyse du compte de résultat

##### Analyse verticale

L’analyse verticale du compte de résultat consiste à rapporter tous les éléments du compte de résultat au chiffre d’affaires. C’est-à-dire la part de chaque élément du compte de résultat dans le processus de création de richesse.

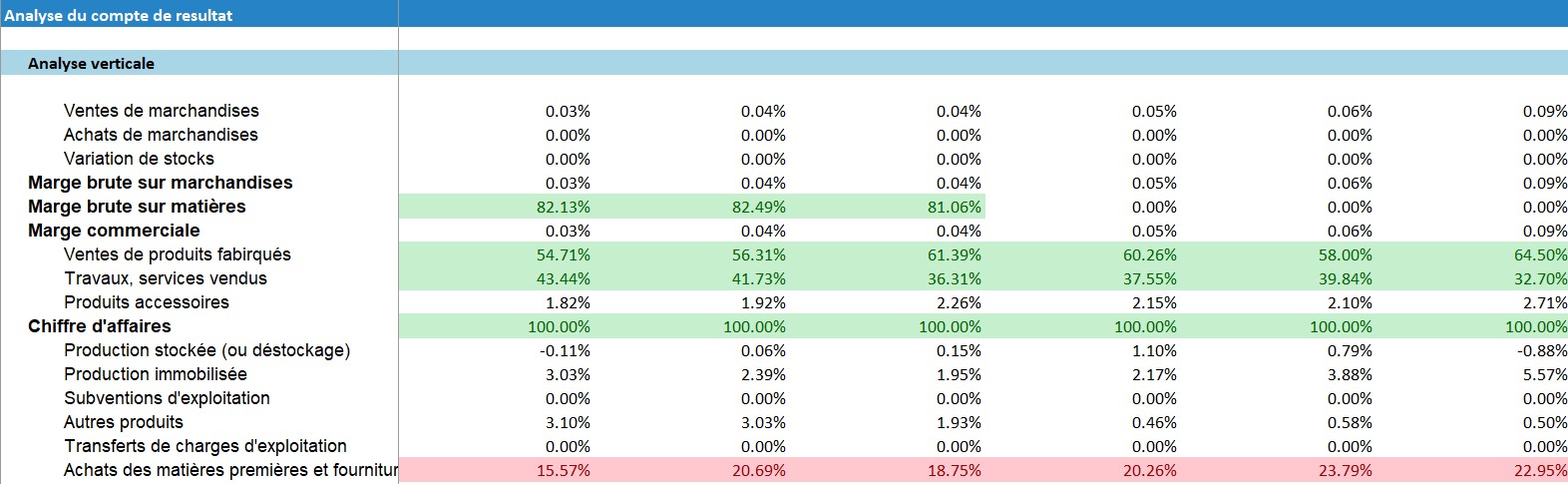
Qu’allons-nous faire exactement ? nous allons tout simplement diviser chaque rubrique du compte de résultat par le chiffre d’affaires et nous obtiendrons notre tableau pour l’analyse verticale.

Figure 2 Analyse verticale du compte d resultat (extrait)

##### Analyse horizontale

Pour ce qui est de l’analyse horizontale du compte de résultat, et d’ailleurs cela sera pareil pour le bilan et le tableau des flux de trésorerie, nous allons tous simplement calculer la croissance de chaque élément du compte de résultat.

Équation 27 Taux de croissance 1

Équation 28 Taux de croissance 2

Ce calcul de croissance va nous permettre de voir l’évolution de chaque élément d’une année à la suivante.

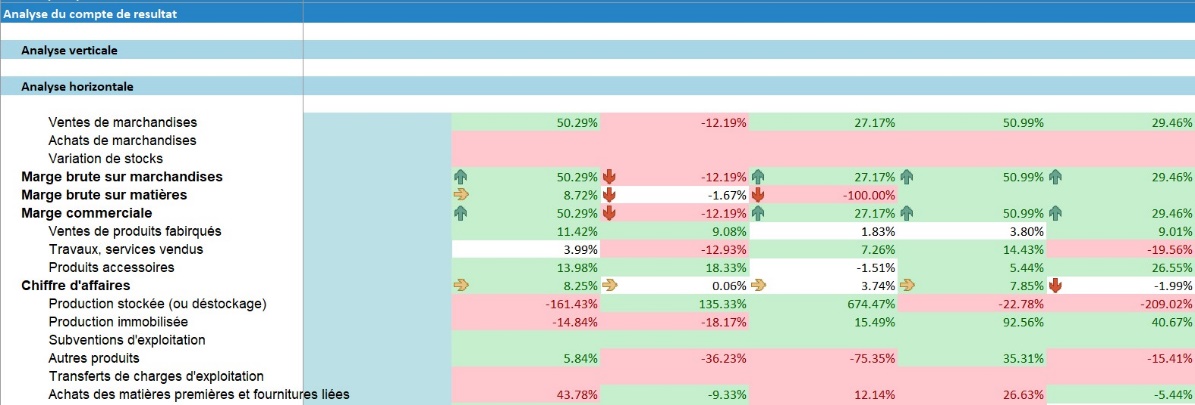


Figure 3 Analyse horizontale du compte d resultat (extrait)

#### Analyse du bilan

##### Analyse verticale

Pour ce qui est de l’analyse verticale du bilan, elle va ressembler un tantinet à celle du compte de résultat. Sauf qu’ici, il y aura deux choses. Le premier c’est rapporter tous les éléments qui contribuent à l’actif au total actif, le second c’est de rapporter tous les éléments qui contribuent au passif au total passif.

De ce fait, nous nous retrouvons avec un tableau ou nous allons voir apparaitre le total actif et la part des éléments qui le composent et de même pour le passif.

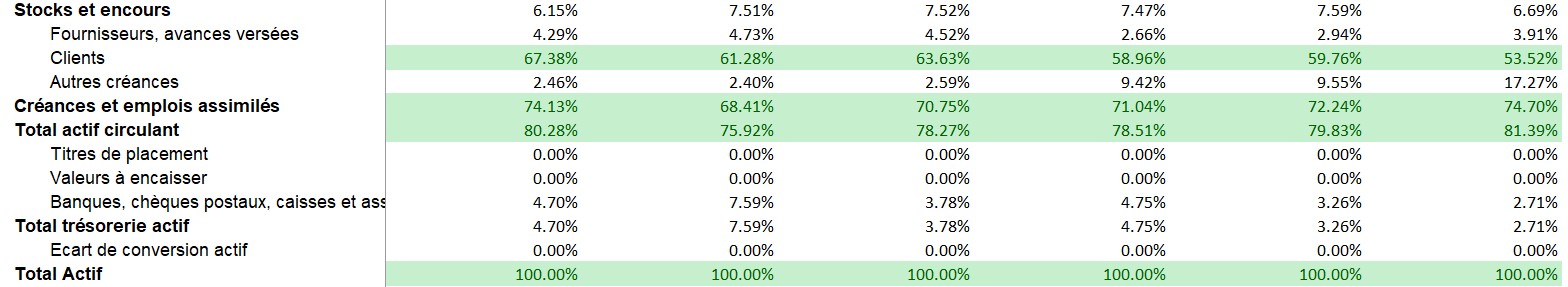


Figure 4 Analyse verticale du bilan : actifs (extrait)

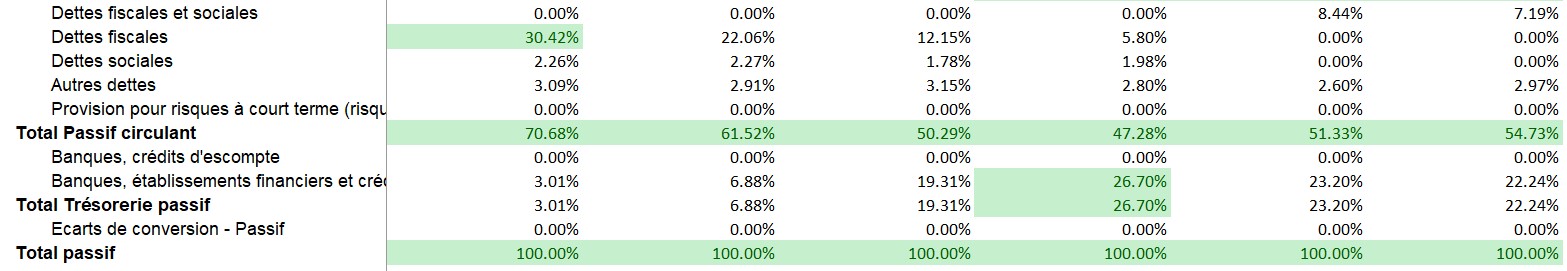


Figure 5 Analyse verticale du bilan : passifs (extrait)

##### Analyse horizontale

L’analyse horizontale du bilan ne sera en rien différente de celle du compte de résultat comme on l’avait dit. Nous allons calculer la croissance de chaque élément, et plus haut la formule de la croissance est présentée.

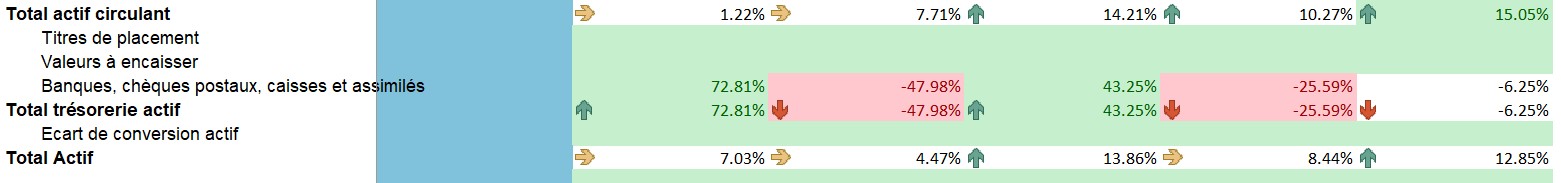


Figure 6 Analyse horizontale du bilan : actifs (extrait)

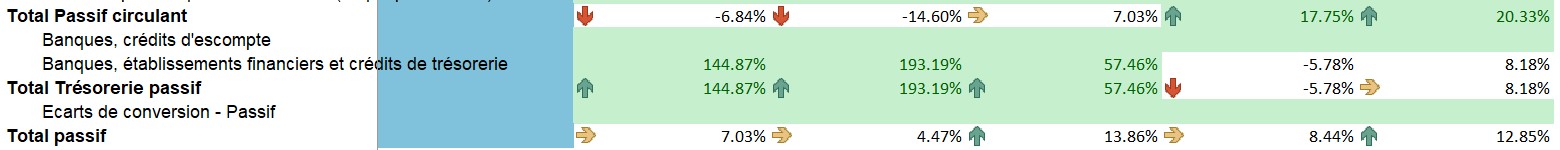


Figure 7 Analyse horizontale du bilan : passifs (extrait)

#### Analyse du tableau des flux de trésorerie

##### Analyse verticale

Pour ce qui est de l’analyse du tableau des flux de trésorerie, l’élément de base va être la trésorerie au 31 décembre, c’est la trésorerie finale. Et après on fait la même chose que les autres états financiers, ce qui veut dire calculer le rapport de tous les éléments sur la trésorerie finale.

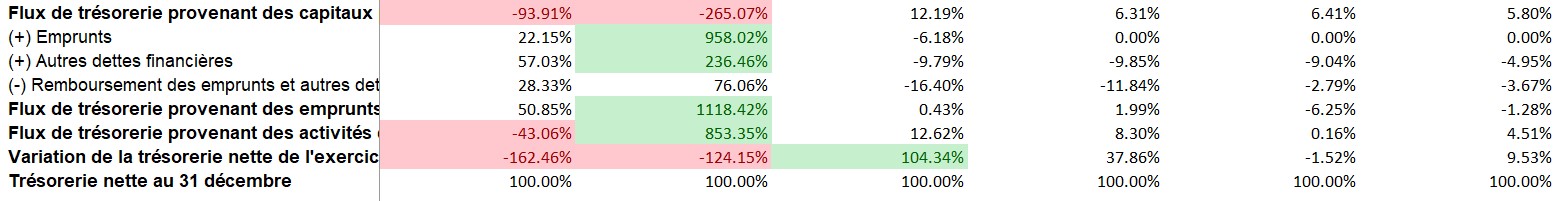


Figure 8 Analyse verticale du tableau des flux de tresorerie (extrait)

##### Analyse horizontale

Nous l’avons fait et refaits pour le compte de résultat et le bilan, et nous allons le refaire pour le tableau des flux de trésorerie, par conséquent calculer la croissance de chaque élément.

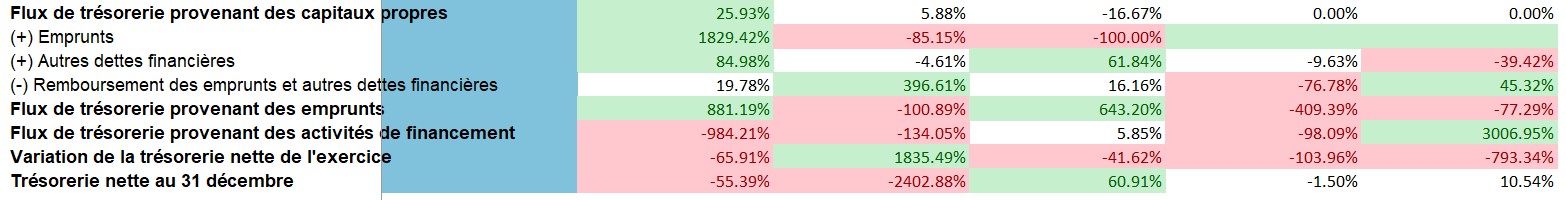


Figure 9 Analyse horizontale du tableau des flux de trésorerie (extrait)

### Analyse de l’activité et des relations de trésorerie

Ce point de l’analyse financière est la partie qui va nous permettre d’apprécier l’évolution des éléments clés d’une entreprise. C’est tous les chiffres que l’entreprise doit maximiser pour rester en bonne santé et durée dans le temps. Il y aura trois (3) d’analyse qui va être faite dans cette partie.

#### Analyse du cycle de vie de l'activité

Le cycle de vie de l’activité peut être jugée en analysant trois éléments fondamentaux des états financiers. A savoir, d’abord le chiffre d’affaires qui représente la création de richesses. Ensuite, le résultat net qui est ce qui reste pour l’entreprise après s’être acquitté de toutes ses charges. Enfin la trésorerie nette qui correspond à l’argent effectivement disponible (caisse ou banque).

Il faut noter que durant cette analyse on ne va pas calculer, l’analyse du cycle de vie va regarder l’évolution de ses valeurs et se faire une opinion du cycle d’activité de l’entreprise.

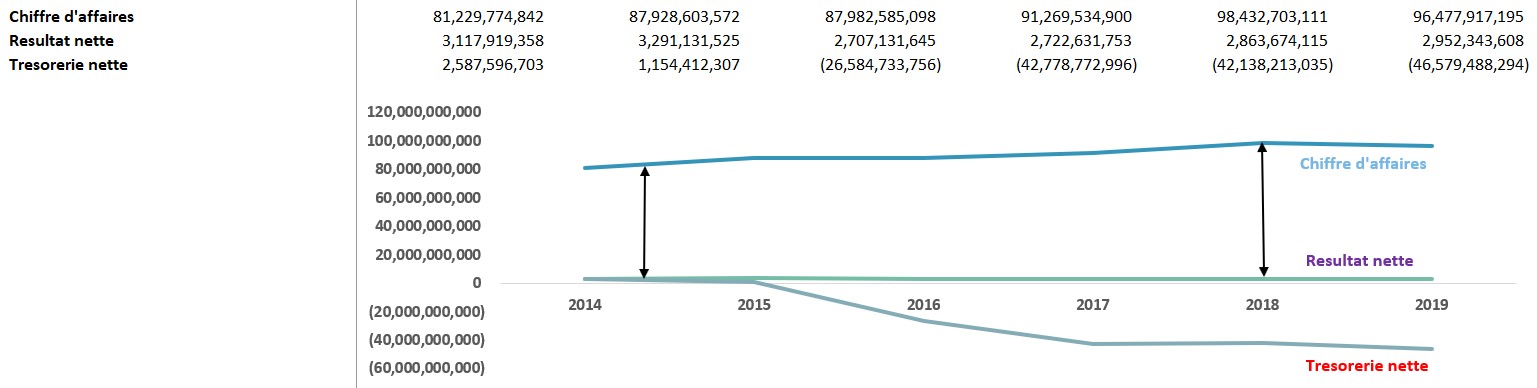


Figure 10 Graphique de l'analyse du cycle de vie de l'activité

#### Analyse du comportement des flux de trésorerie

Comme fait précédemment, cette analyse va nous permettre de constater l’évolution des rubriques les plus importants du tableau des flux de trésorerie. Ici, l’analyse va voir et apprécier les flux de trésorerie liés aux activités opérationnelles, les flux de trésorerie liés aux activités d’investissement, les flux de trésorerie liés aux activités de financement, la variation de la trésorerie nette et enfin la trésorerie finale.

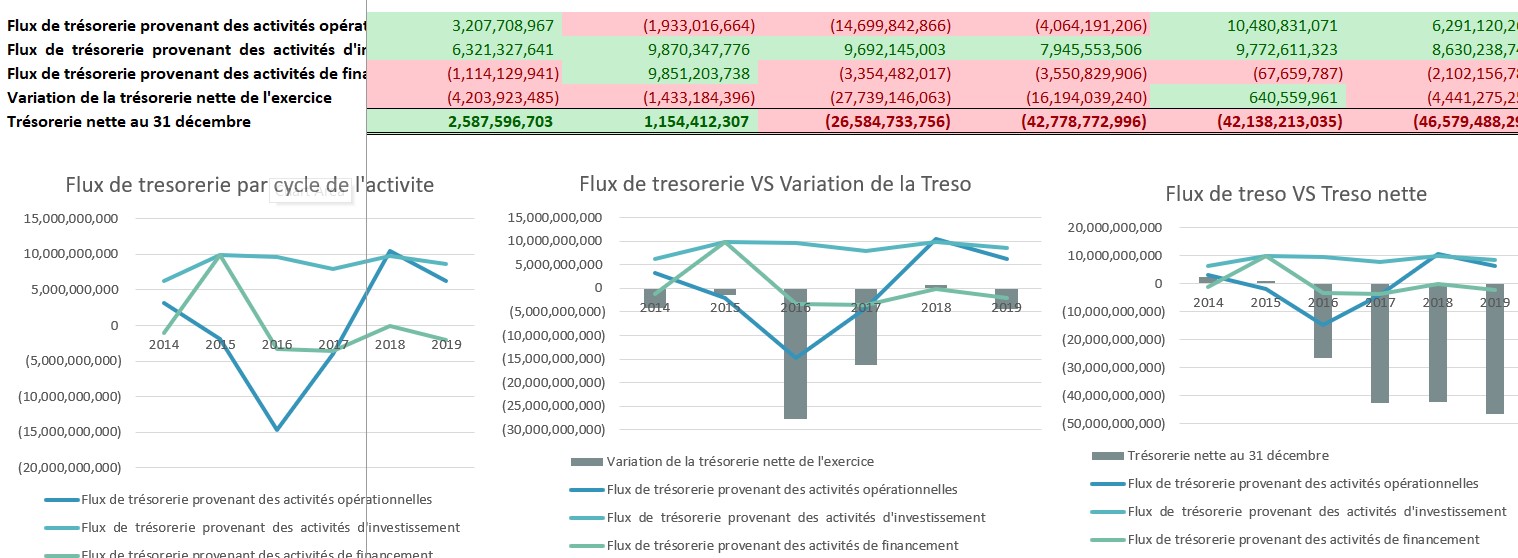


Figure 11 Graphique de l'analyse du comportement des flux de trésorerie

#### Analyse des équilibres financiers et la relation de trésorerie

Les équilibres financiers sont composés de trois sous-partie et chacune d’entre elles va relever un tableau qui va nous donner encore plus d’éclaircissement sur les finances de l’entreprise. Nous avons entre autres le bilan économique, les équilibres financiers, la relation de trésorerie.

##### Bilan économique

Le bilan économique est un mini-bilan qui va faire ressortir les avoirs et dus de l’entreprise. Nous pouvons notamment y retrouver les actifs immobilisés, les capitaux, les endettements… Cependant, le rôle principal du bilan économique, c’est de calculer l’actif économique et les capitaux investis. Il faut ajouter que ce mini-bilan est bien fait si l’actif économique est égal aux capitaux investis.

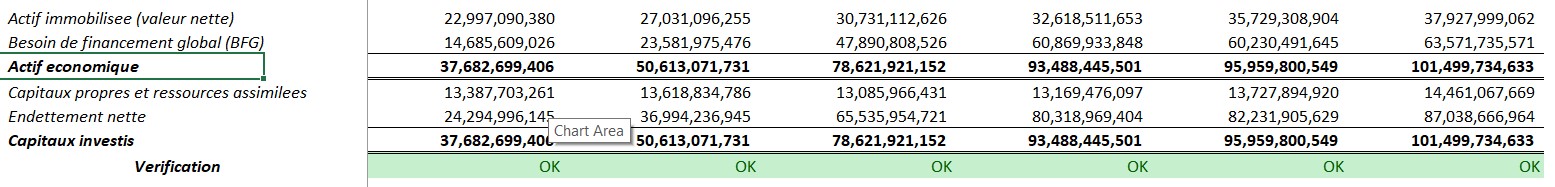


Figure 12 Le bilan économique

##### Les équilibres financiers

L’équilibre n’est plus, ni moins que la trésorerie nette recalculée et vérifiée si c’est conforme à ce qui est aux valeurs présentées au niveau du tableau des flux de trésorerie. Cette trésorerie est obtenue en faisant la différence entre le fonds de roulement nette global et le besoin de financement global.

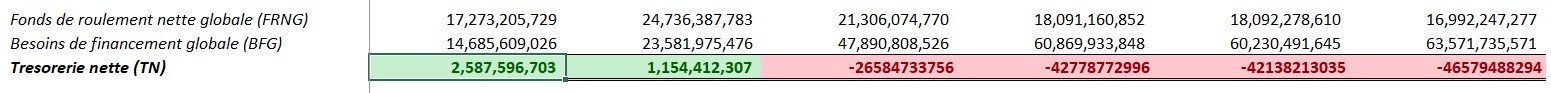


Figure 13 Les équilibres financiers

##### La relation de trésorerie

Pour ce qui est de la relation de trésorerie, son but est d’apprécier la création de richesse, les différents besoins de l’entreprise liée à la trésorerie (fonds de roulement et financement) et l’argent qui est effectivement à la disposition de l’entreprise. Eu égard à cela, il est logique de voir apparaitre dans son tableau le chiffres d’affaires, le besoin de fonds de roulement global, le besoin de financement global et la trésorerie nette.

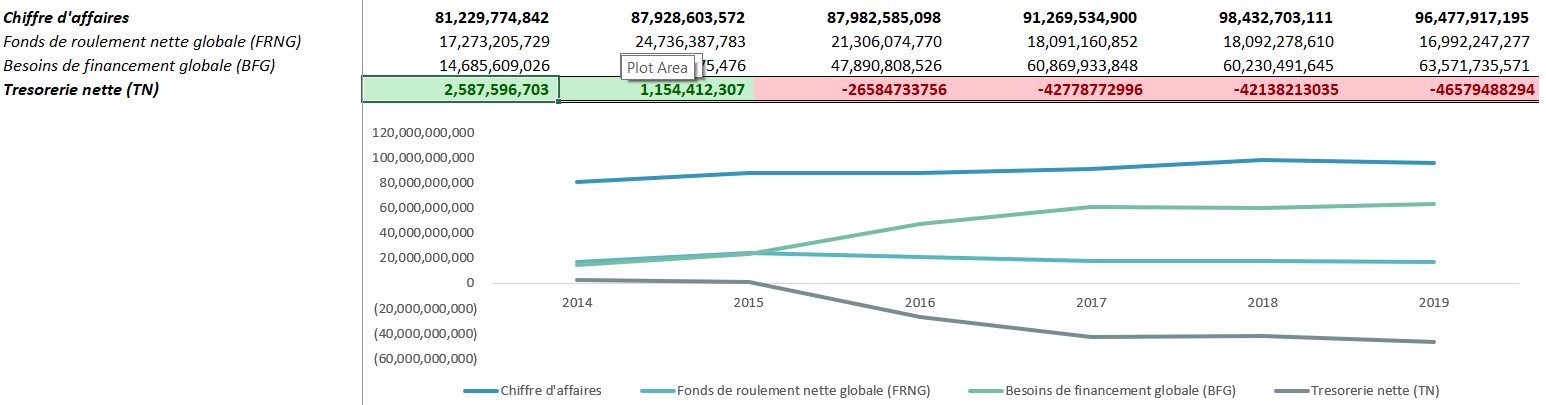


Figure 14 La relation de trésorerie

#### Profitabilité

C’est la famille de ratio de profitabilité pour la plupart d’entre eux calculer à partir du compte de résultat. Cela nous permet de constater la part des SIG dans la création de richesse. C’est ainsi que nous allons tous voir leur part sur le CA.

##### Calcul des ratios de profitabilité

Tableau 4 Les ratios de profitabilité

|  |  |
| --- | --- |
| Ratio | Formule |
| Taux de marge commerciale |  |
| Taux de valeur ajoutée |  |
| Taux bruts d’exploitation |  |
| Taux marge d’exploitation |  |
| Taux de marge nette |  |
| Taux de performance opérationnelle |  |

Tableau 5 Calcul des ratios de profitabilité

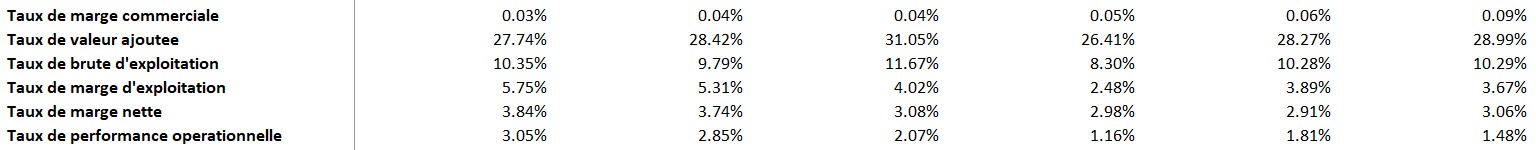


Figure 15 Calcul des ratios de profitabilité

##### Analyse tendancielle des ratios de profitabilité

Une analyse tendancielle est la comparaison des ratios sur toute la période d’étude par rapport à une année, cette année peut être n’importe laquelle (généralement une année qui est marquée par un évènement fort), mais ici nous choisissons la première par souci de simplicité.

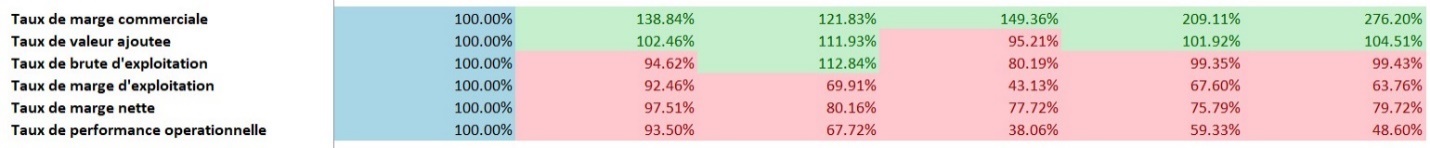


Figure 16 Analyse tendancielle des ratios de profitabilité

##### Représentation graphique des ratios de profitabilité

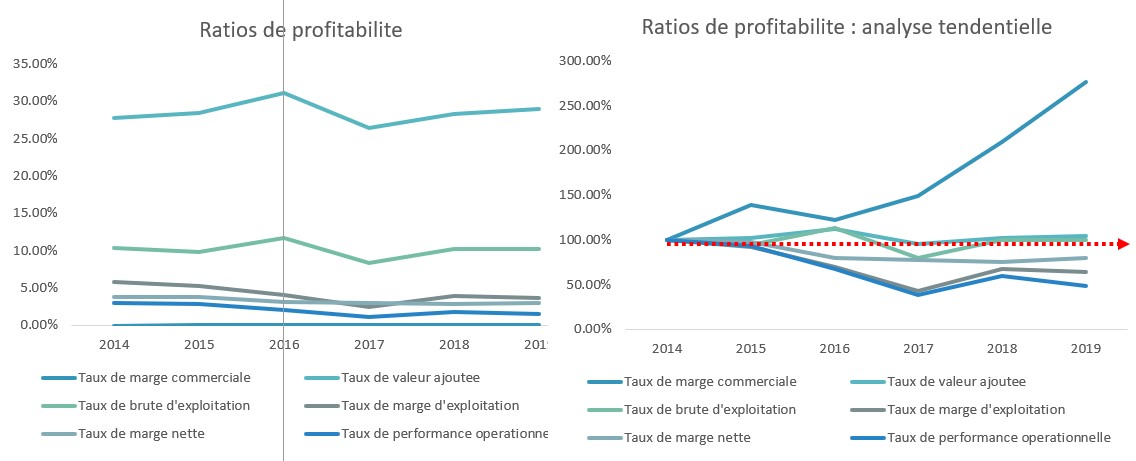


Figure 17 Représentation graphique des ratios de profitabilité

#### Rentabilité

Comme son nom l’indique cette famille se veut être une mesure de la rentabilité. Il y en a deux types : la rentabilité financière et la rentabilité économique. C’est une famille qui suscite non seulement l’intérêt de l’entreprise et ses actionnaires mais aussi de potentiels investisseurs.

##### Calcul des ratios de rentabilité

Tableau 6 Les ratios de rentabilité

|  |  |
| --- | --- |
| Ratios | Formule |
| **Rentabilité économique** | |
| Méthode directe |  |
| Taux de marge d’exploitation (1) |  |
| Taux de rotation de l’actif économique (2) |  |
| Méthode indirecte |  |
| **Rentabilité financière** | |
| Méthode directe |  |
| Taux de marge nette (3) |  |
| Taux de rotation de l’actif économique (4) |  |
| Facteur d’ajustement (5) |  |
| Méthode indirecte |  |

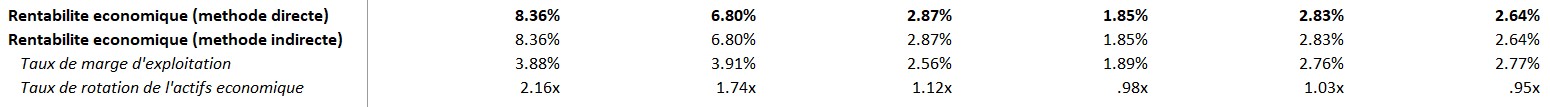


Figure 18 Calcul des ratios de rentabilité

##### Analyse tendancielle des ratios de rentabilité

* Rentabilité économique



Figure 19 Analyse tendancielle des ratios de rentabilité economique

* Rentabilité financière

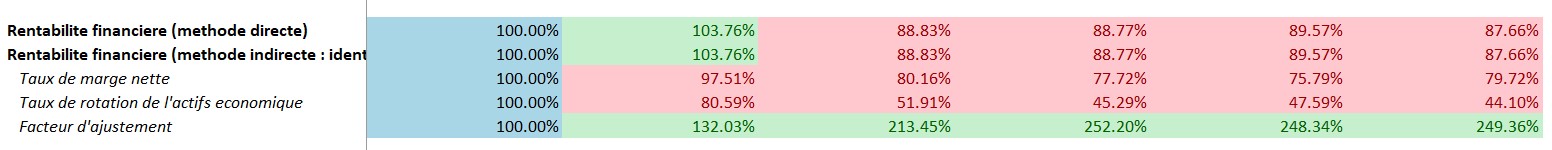


Figure 20 Analyse tendancielle des ratios de rentabilité financière

##### Représentation graphique des ratios de rentabilité

* Rentabilité économique

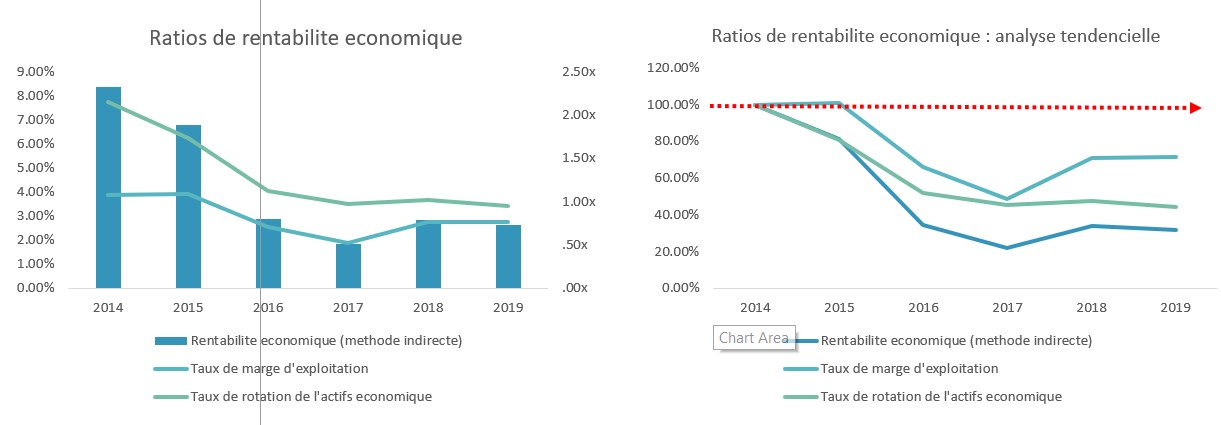


Figure 21 Représentation graphique des ratios de rentabilité economique

* Rentabilité financière

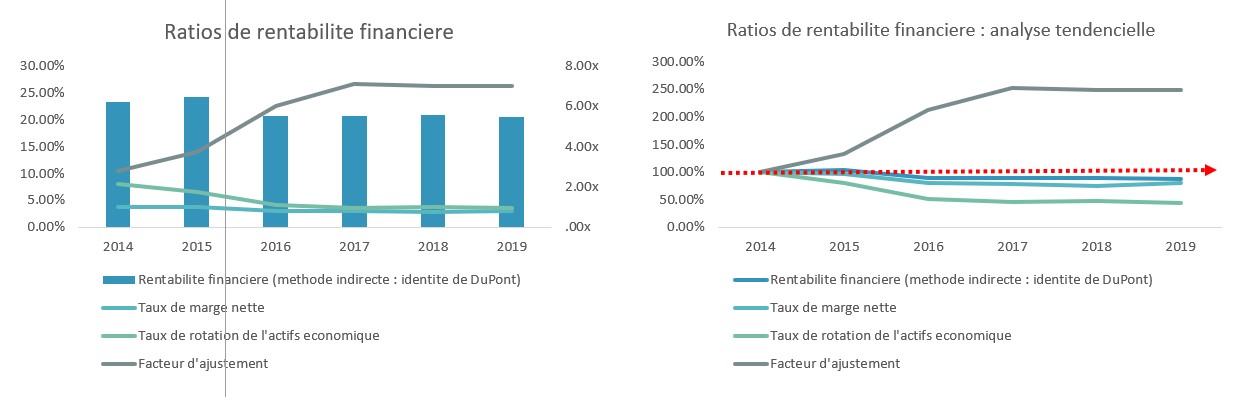


Figure 22 Représentation graphique des ratios de rentabilité financière

#### Politique comptable

La politique comptable est une famille de ratios qui nous permet d’observer l’état des immobilisations. Apres calcul, on peut se rendre compte de comment sont amortis les immobilisations et quel politique l’entreprise adopte à cet égard.

##### Calcul des ratios de politique comptable

Tableau 7 Les ratios de politique comptable

|  |  |
| --- | --- |
| Ratios | Formule |
| Ratio de vétusté |  |
| Taux de provision de stocks |  |
| Taux de provision de créances |  |

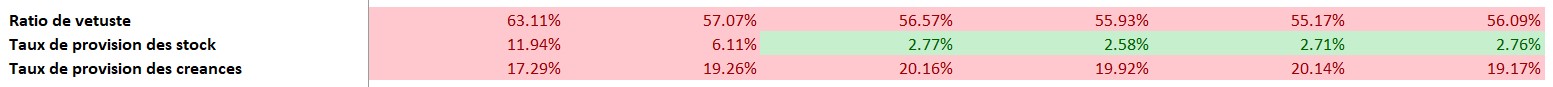


Figure 23 Calcul des ratios de politique comptable

##### Analyse tendancielle des ratios de politique comptable

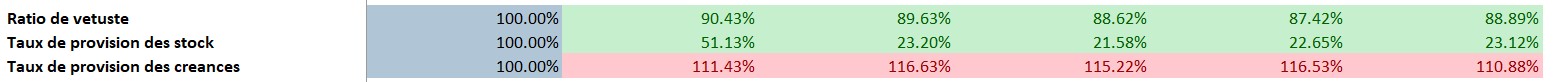


Figure 24 Analyse tendancielle des ratios de politique comptable

##### Représentation graphique des ratios de politique comptable

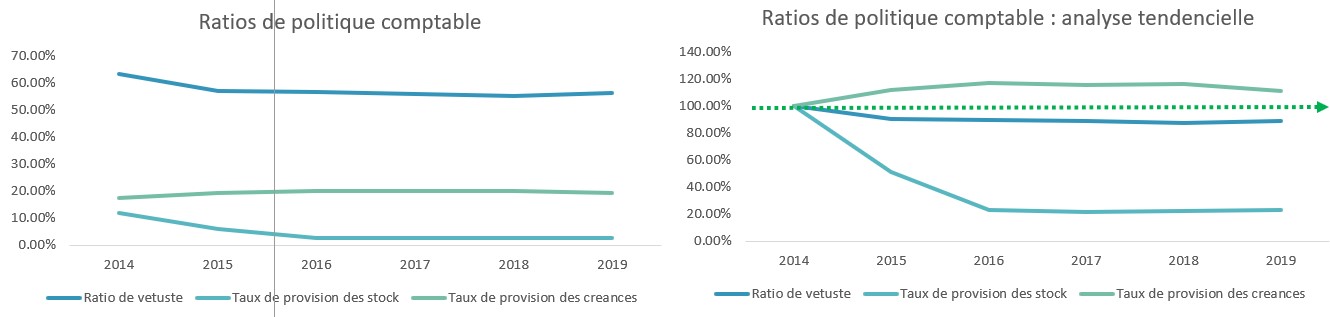


Figure 25 Représentation graphique des ratios de politique comptable

#### Liquidité

La liquidité est l’aptitude qu’a l’entreprise à faire face à ses dépenses par la circulation optimale des flux de trésorerie. Cette famille de ratios va nous permettre de voir comment l’entreprise est liquide sous les trois ratios de liquidité.Calcul des ratios de liquidité.

##### Calcul des ratios de liquidité

Tableau 8 Les ratios de liquidité

|  |  |
| --- | --- |
| Ratios | Formule |
| Liquidité générale |  |
| Liquidité réduite |  |
| Liquidité immédiate |  |

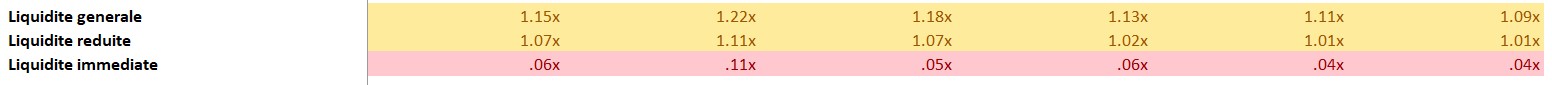


Figure 26 ratios de liquidité

##### Analyse tendancielle des ratios de liquidité

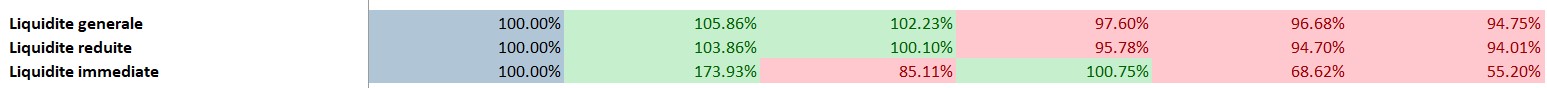


Figure 27 Analyse tendancielle des ratios de liquidité

##### Représentation graphique des ratios de liquidité

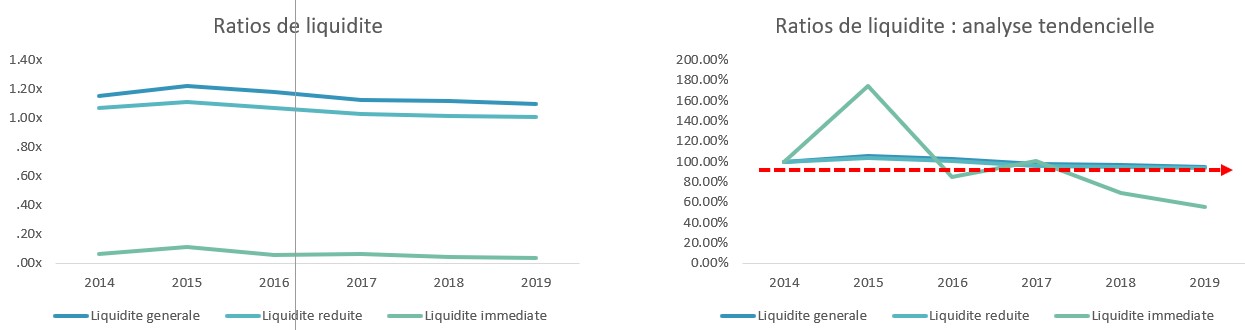


Figure 28 Représentation graphique des ratios de liquidité

#### Gestion de la dette

La famille de ratios gestion de la dette est très facile à comprendre, puisque son nom en dit tout. Ici, nous allons calculer quelques ratios qui vont nous permettre d’analyser comment endettée l’entreprise est et quelle politique de remboursement elle adapte.

##### Calcul des ratios de gestion de la dette

Tableau 9 Les ratios de gestion de la dette

|  |  |
| --- | --- |
| Ratios | Formule |
| Levier |  |
|  |
| Taux d’endettement |  |
| Maturité de l’endettement |  |
| Couverture des frais financiers |  |
| Cout implicite de la dette |  |
| Capacite de remboursement |  |

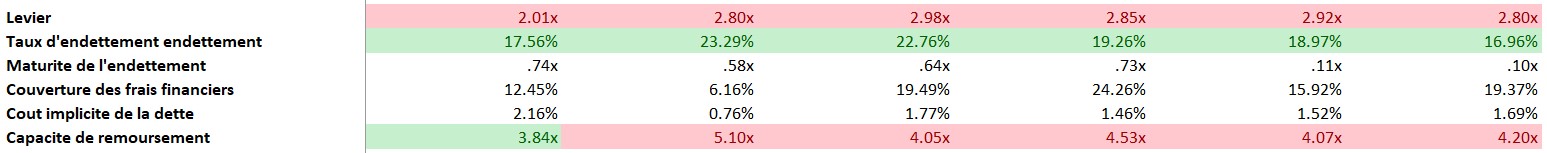


Figure 29 Calcul des ratios de gestion de la dette

##### Analyse tendancielle des ratios de gestion de la dette

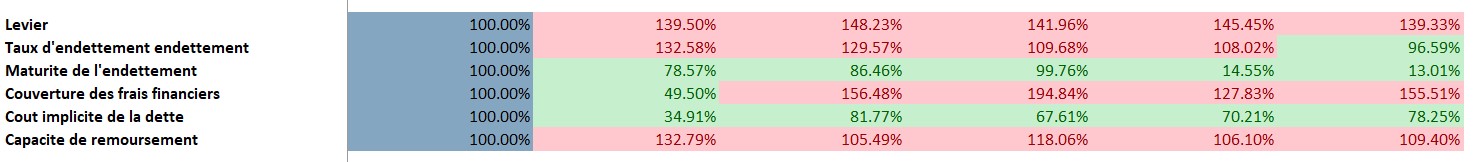


Figure 30 Analyse tendancielle des ratios de gestion de la dette

##### Représentation graphique des ratios de gestion de la dette

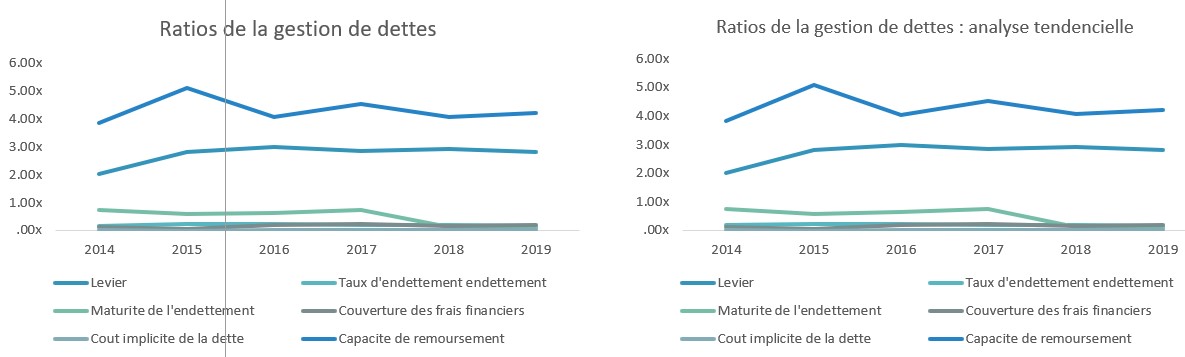


Figure 31 Représentation graphique des ratios de gestion de la dette

#### Flux de trésorerie

La famille de ratios des flux de trésorerie nous informe principalement sur les stratégies d’investissement adoptées par l’entreprise et leur efficacité.

##### Calcul des ratios de flux de trésorerie

Tableau 10 Les ratios de flux de trésorerie

|  |  |
| --- | --- |
| Ratios | Formule |
| Capacité à investir |  |
| Taux d’investissement net |  |
| Taux de réinvestissement |  |
| Taux de croissance |  |
| Couverture des flux de trésorerie des activités opérationnelles |  |

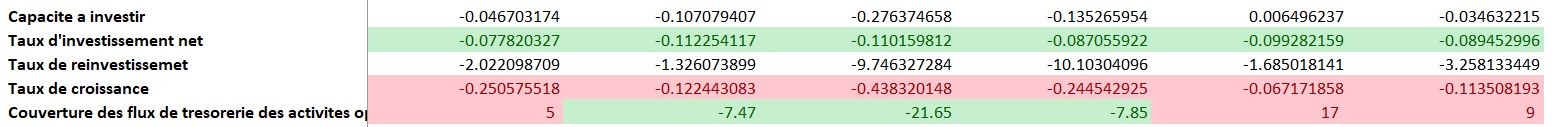


Figure 32 Calcul des ratios de flux de trésorerie

##### Analyse tendancielle des ratios de flux de trésorerie

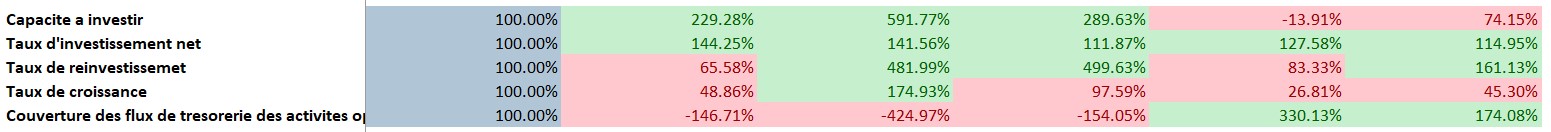


Figure 33 Analyse tendancielle des ratios de flux de trésorerie

##### Représentation graphique des ratios de flux de trésorerie

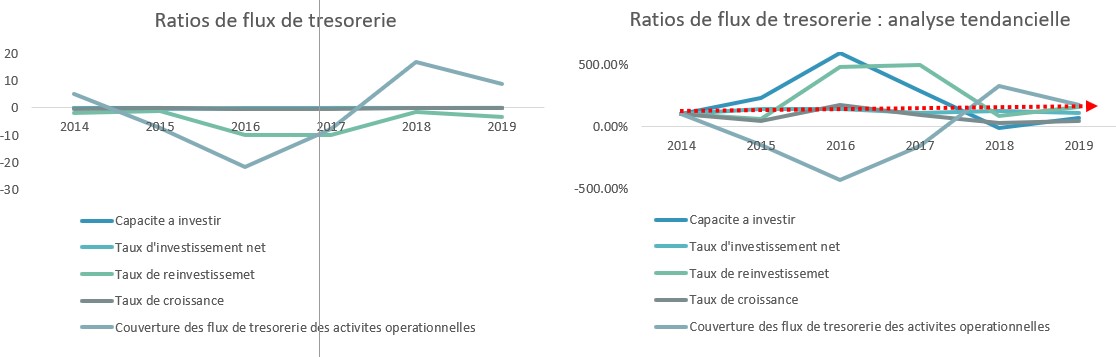


Figure 34 Représentation graphique des ratios de flux de trésorerie

#### Efficacité des actifs du BFG

C’est une famille qui regroupe les ratios sur l’efficacité du BFG. Nous allons jongler entre le compte de résultat et le bilan pour faire les calculs.

##### Calcul des ratios d’efficacité des actifs du BFG

Tableau 11 Les ratios d’efficacité des actifs du BFG

|  |  |
| --- | --- |
| Ratios | Formule |
| Variation du chiffre d’affaires |  |
| Variation de la valeur ajoutée |  |
| Taux de valeur ajoutée |  |
| Efficience du personnel |  |
|  |
|  |
| Besoin de financement global |  |
|  |
| Besoin de financement global en jours |  |
| Politique d’investissement (PI) |  |
|  |
| Politique de financement (PF) |  |
|  |
| Rotation de l’actif (RA) |  |
| Rotation de l’actif économique (RAE) |  |
| Délai de rotation des stocks (DRS) |  |
|  |
| Délai de récupération des créances clients (DRCC) |  |
| Délai de paiement aux fournisseurs (DPF) |  |
| **Cycle d’exploitation (CE)** |  |
| **Cycle de trésorerie (CT)** |  |



Figure 35 Calcul des ratios d’efficacité des actifs du BFG

##### Analyse tendancielle des ratios d’efficacité des actifs de BFG

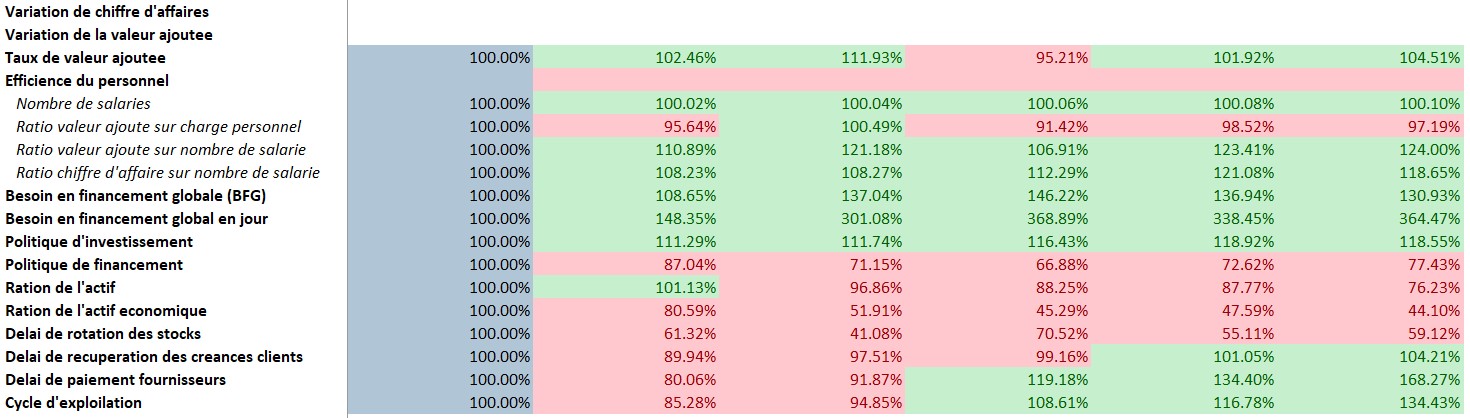


Figure 36 Analyse tendancielle des ratios d’efficacité des actifs de BFG

##### Représentation graphique des ratios d’efficacité des actifs de BFG

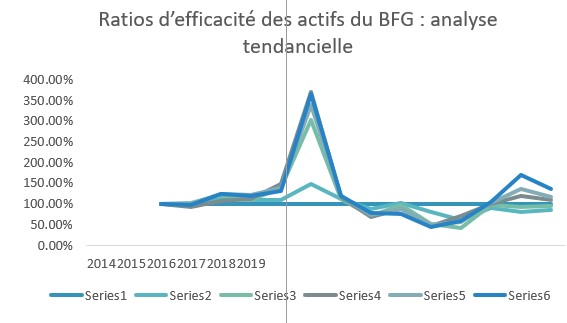


Figure 37 Représentation graphique des ratios d’efficacité des actifs de BFG

#### Valorisation

Nous allons terminer les familles de ratios par celle de la valorisation. Mais dans ce cas, nous allons nous concentrer sur la valorisation boursière. Nous serons en mesure d’observer comment l’entreprise se comporte sur les marchés financiers, la bourse régionale des valeurs mobilières (BRVM) en l’occurrence.

##### Calcul des ratios de valorisation

Tableau 12 Les ratios de valorisation

|  |  |
| --- | --- |
| Ratios | Formule |
| Capitalisation des bénéfices |  |
|  |
| Valorisation boursière |  |
|  |
| Capitalisation du chiffre d’affaires |  |
| Valeur de marché de l’actif économique |  |

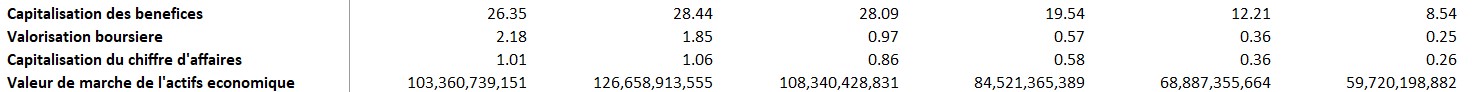


Figure 38 Calcul des ratios de valorisation

##### Analyse tendancielle des ratios de valorisation

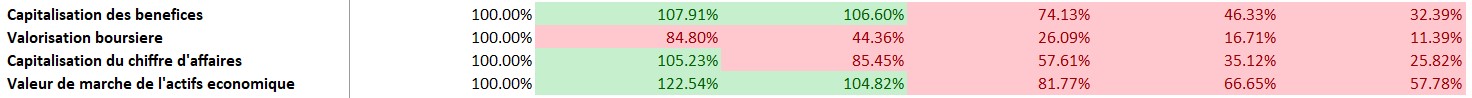


Figure 39 Analyse tendancielle des ratios de valorisation

##### Représentation graphique des ratios de valorisation

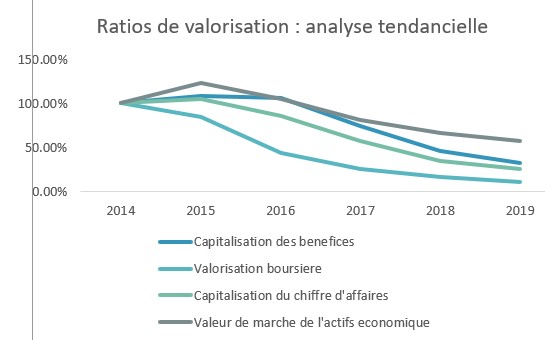


Figure 40 Représentation graphique des ratios de valorisation

#### Analyse de la probabilité de défauts

L’analyse de probabilité de défaut, aussi appelé risque de défaut représente des méthodes pour l’analyste financier de prédire, en lisant des constantes de l’entreprise, si cette dernière va être en défaut, voir la faillite. Il est possible de prédire si une entreprise va être en défaut en utilisant différentes méthodes.

Comme déjà dit, plusieurs chercheurs et structures ont proposé des méthodes évaluation du risque de défauts d’une entreprise. Ces méthodes reposent sur des calculs mathématiques et des interprétations qui selon le cas de figure vont nous permettre de savoir si l’entreprise est un défaut ou pas. En plus des valeurs provenant des états financiers, nous aurons besoin des informations de l’entrepris issues des marchés financiers. Car oui, Cette analyse va se faire seulement pour les entreprises cotées en bourse.

Pour ce document, nous allons voir la méthode proposée par Edward I. Altman en 2005. Cette méthode se veut être une mesure de la probabilité de défauts pour les marchés émergents. Edward I. Altman a développé d’autres méthodes mais c’est celle de 2005 qui va nous intéresser.

##### Quelques informations boursières

Avant de commencer quelconque analyse, nous aurons besoin d’informations ne se trouvant pas dans les états financiers. Il va falloir d’abord chercher les informations boursières relatives à cette entreprise. Les données boursières nécessaire pour l’analyse de probabilité de défauts sont :

* Le nombre d’actions
* Le cours boursier (prix d’une action)
* La capitalisation boursière

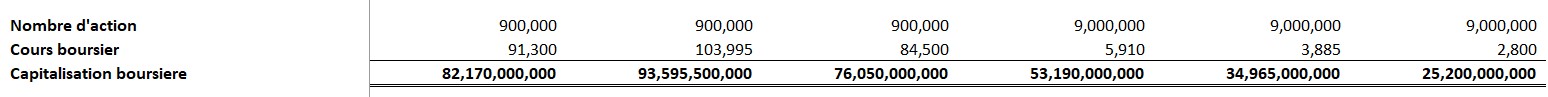


Figure 41 Quelques informations boursières

##### Calcul du Zscore de Altman

La méthode de Altman est un moyen qui va nous permettre de calculer ce qu’il appelle zscore, ce qui aura pour objectif de nous renseigner sur deux choses : la zone et le rating. D’abord, il faut dire il y a trois zones avec cette méthode, zone de sécurité, zone d’incertitude et la zone de détresse. Pour ce qui est du rating, elle va de AAA à D ; AAA étant la plus forte bien évidemment.

A cet effet, nous devons bien entendu calculer le zscore, mais comment ? Le zscore est la sommation de certains ratios (nous sommes toujours dans d’analyse par les ratios), et de leurs poids associés. Les poids ont déjà été calculer par Altman nous avons à juste les utiliser. Le tableau suivant va résumer les ratios et leur poids.

Tableau 13 Les ratios du Zscore d'Altman

|  |  |
| --- | --- |
| Ratios | Poids |
|  | 1 |
|  | 6,56 |
|  | 3,26 |
|  | 6,72 |
|  | 1.05 |

Par conséquent la formule du zscore est :

Équation 29 Zscore d'Altman

Xi représente les ratios

Wi représente les poids.

Maintenant après avoir calculer le zscore, nous devons apprécier l’implication, c’est-à-dire juger de la zone et du rating.

Tableau 14 Le rating et les zones du zscore d'Altman

|  |  |  |
| --- | --- | --- |
| Zscore | Rating | Zone |
| > 8,15 | AAA | Zone de sécurité |
| 7,60 - 8,15 | AA+ |
| 7,30 - 7,60 | AA |
| 7,00 - 7,30 | AA- |
| 6,85 - 7,00 | A+ |
| 6,65 - 6,85 | A |
| 6,40 - 6,65 | A- |
| 6,25 - 6,40 | BBB+ |
| 5,85 - 6,25 | BBB |
| 5,65 - 5,85 | BBB- | Zone d’incertitude |
| 5,25 - 5,65 | BB+ |
| 4,95 - 5,25 | BB |
| 4,75 - 4,95 | BB- |
| 4,50 - 4,75 | B+ |
| 4,15 - 4,50 | B |
| 3,75 - 4,15 | B- | Zone de détresse |
| 3,20 - 3,75 | CCC+ |
| 2,50 - 3,20 | CCC |
| 1,75 - 2,50 | CCC- |
| < 1,75 | D |

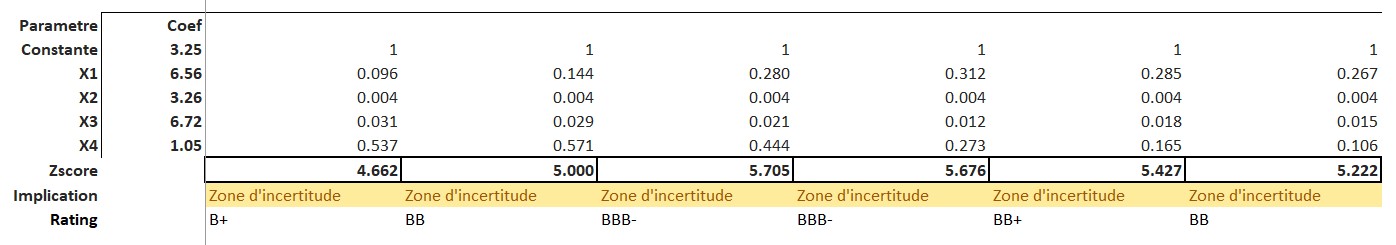


Figure 42 Calcul du Zscore d'Altman

### Evaluer l’entreprise

Eu égard à tout ce qui a été dit et fait plus haut, on peut donner un avis général sur l’entreprise à travers son évaluation. L’évaluation de l’entreprise est l’exercice par lequel l’analyste va donner son opinion le plus objectif possible sur la santé financière de l’entreprise. Cette opinion va être naturellement basée sur les résultats de l’analyse financière mais pas que. Oui, trop souvent il y a des facteurs macroéconomiques qui entrent en jeu dans le processus de création de richesse d’une entreprise. L’analyste financier doit être en mesure de trouver cela et le prendre en compte dans l’évaluation de l’entreprise.

Durant la crise des surprimes, beaucoup d’entreprises avaient des résultats douteux mais ce n’est pas pour cela que l’on accordait une mauvaise note a cette dernière.

Les enjeux macroéconomiques peuvent s’avérer être déterminants, surtout dans le monde d’échange où nous vivons. Donc cet aspect doit toujours attiser l’attention de l’analyste avant d’émettre un quelconque jugement sur une entreprise.

Ceci étant dit, cette évaluation faite à travers un rapport que l’analyste va rédiger et qui va ensuite être soumis au dirigeant, aux investisseurs, ou tout autres personnes physique ou morale s’intéressant à la structure analysée. L’analyste va reprendre les rubriques de l’analyse, l’explique d’une manière claire et concise à ce que des non financiers puissent le comprendre.

* Annexe 2 : interprétation SIG
* Marge commerciale

Après le calcul de la marge commerciale deux cas de figure se présentent à nous :

* Marge commercial positif : cela veut dire que l’entreprise est totalement rentable car les produits dépassent les charges.
* Marge commercial négatif : dans cette situation l’entreprise est non rentable. Deux solutions se présentent à elle : soit augmenter ses ventes ou réduire ses charges.
* Chiffre d’affaires

Le CA permet de mesurer les ventes de l’entreprise, plus il augmente, plus l’entreprise est bien positionnée sur le marché. On peut ajouter que beaucoup d’autres indicateurs vont dériver du CA, et ce dernier a un impact positif ou négatif sur le résultat de l’exercice. Cependant comme déjà dit, il n’est pas le meilleur des indicateurs, il va falloir le combiner avec les autres SIG, dont nous allons parler plus tard.

* Valeur ajoutée
* Valeur ajoutée faible : l’entreprise créée peu de richesse, elle ne vend pas assez ou subit beaucoup trop de charges. Cette entreprise doit les gérer si elle n’a pas envie de perdre la confiance des investisseurs.
* Valeur ajoutée élevée : c’est ici que tout le monde a envie d’être, l’entreprise crée assez de richesse, elle pourra faire face aux charges qui restent, notamment les frais de personnel. L’entreprise est en bonne santé et peut attirer les investisseurs.
* Résultat d’exploitation
* REX positif : l’entreprise est rentable grâce à tous les éléments qui composent son cycle d’exploitation, l’entreprise dégage un bénéfice d’exploitation.
* REX négatif : l’entreprise doit revoir son modèle économique, car ne dégageant pas de bénéfice sur son cycle d’exploitation.
* L’excèdent brute d’exploitation

L’EBE est la création de richesse en excluant les éléments financiers et exceptionnels :

* EBE positif : l’entreprise crée de la valeur et ses produits couvrent tous ses charges, l’entreprise est dans une très bonne santé et doit continuer sur ce rythme.
* EBE négatif : l’entreprise est dans un risque car n’étant pas du tout rentable, elle doit corriger si elle ne veut pas tomber en faillite.
* Le résultat financier
* RF positif : cela veut dire que l’entreprise génère assez de revenu grâce à ses placements, en plus, l’entreprise a atteint un équilibre financier et aussi nous permet de connaitre la somme versée pour les intérêts.
* RF négatif : même si un résultat financier négatif n’est pas forcément une mauvaise chose, cela peut témoigner d’un défaut de gestion de placement, un déséquilibre financier…
* Le résultat exceptionnel
* REP positif : cela veut les produits exceptionnels sont supérieurs aux charges exceptionnelles et notre résultat net va gonfler. Attention ! il n’est pas pour autant un bon indicateur car varie facilement d’une année à l’autre.
* REP négatif : l’entreprise supporte plus de charges exceptionnelles qu’elle ne génère de produits exceptionnels et donc le résultat net va diminuer, c’est non plus forcément une mauvaise chose.
* Le résultat net
* RN positive : c’est ce que tout entreprise cherche, dans ce cas de figure on parle de bénéfice net, l’entreprise dégage une plus-value sur toutes ses activités. Néanmoins, le résultat net comporte des éléments exceptionnels donc n’est pas considéré par certains comme le meilleur des indicateurs.
* RN négative : si une entreprise a un résultat net négatif on parle plutôt de déficit, et ce déficit va se trouver au bilan dans le report à nouveau. Cette entreprise ne va pas rémunérer ses investisseurs et courent le risque de faillite.
* Annexe 3 : Details calcul algorithme

La regression lineaire

Ci-après un tableau de valeur et sa représentation graphique.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Y | 1.2 | 1.9 | 2.8 | 4.4 | 5.4 | 5.8 | 6.7 | 8.4 | 8.9 | 8.8 |

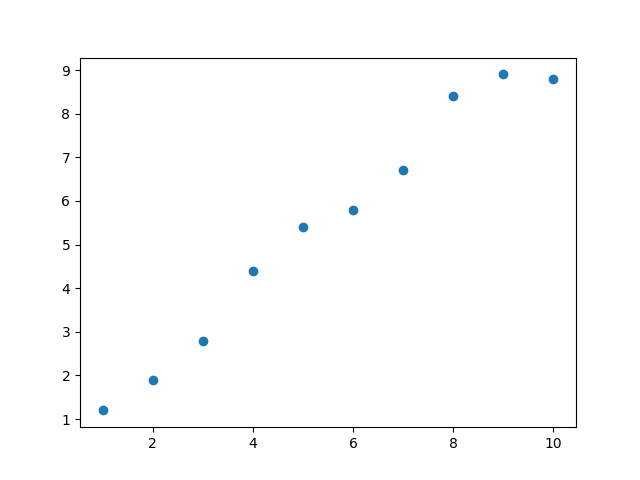


Figure 3 Représentation de points

Prenons cette courbe, ci-dessus nous voyons la représentation d’un certain nombre de points.

Nous pouvons maintenant faire une application de cette fonction avec le premier essai.

Attention ! il ne faut pas oublier que de la même manière que l’on a calculer pour x = 1, il faut aussi le calculer pour tous les autres x et ainsi avoir toutes les erreurs pour pouvoir appliquer la formule générale.

Nous avons trouvé 3.6 il reste qu’a le multiplier avec le Learning rate et w0 est prêt à être mise à jour. Il va falloir faire de même pour w1 et c’est le premier epochs (terme anglais qui signifie le parcours de tous les ligne du jeu de données).

En revanche, le travail ne s’arrête pas là, toutes ses trois actions précédentes, il va falloir les répéter autant de fois que nécessaire pour avoir le modèle le plus fiable possible, généralement on parle de milliers d’epochs. Si le travail est bien fait nous pouvons nous retrouver avec une courbe comme la suivante :

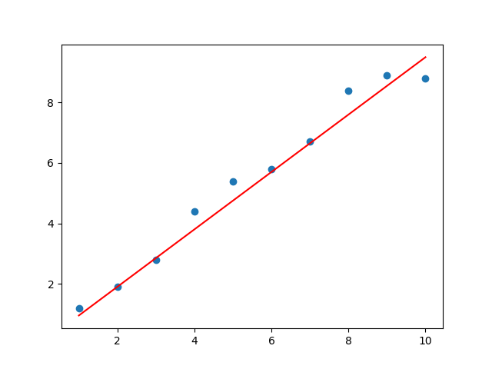


Figure 5 Régression linéaire appliquée à des points

Régression logistique

Nous allons pour la suite faire l’exemple de la fonction logique OÙ :

|  |  |  |
| --- | --- | --- |
| X1 | X2 | OU |
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

Les mêmes étapes, que la régression linéaire, vont revenir avec intronisation d’un nouveau concept : la fonction d’activation (hautement important).

* Forward propogation

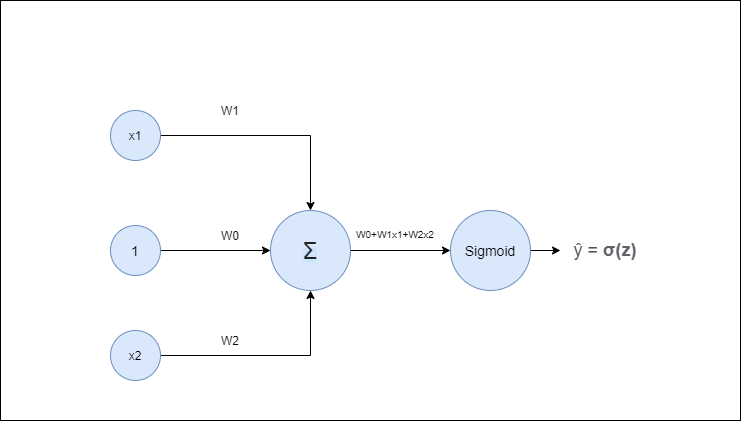


Figure 6 Simple reseau de neurone (Kumar, 2020)

Voici à quoi va ressembler notre réseau de neurones, on va ajouter un autre input en plus x1 et x2, c’est le biais qui va toujours être égale à 1, son utilité est d’éviter que certains neurones ne meurent durant l’entrainement si x1 = 0 et x2 = 0.

Commençons par initialiser

Équation 32 Calcul de la sortie observee

Puisqu’on dit que les valeurs de sortie doivent être 0 ou 1, nous devons trouver un moyen de toujours mettre à l’échelle la sortie observée, c’est là qu’intervient la fonction d’activation. Pour les problèmes de régression logistique il y en a deux très populaires : la fonction à seuil et sigmoid.

|  |  |  |
| --- | --- | --- |
| Fonction | Formule | Sortie possible |
| Seuil(x) |  | 0, 1 |
| Sigmoid (x) | Équation La fonction sigmoid | Tout réel compris en 0 et 1 |

Nous allons continuer avec la fonction sigmoid :

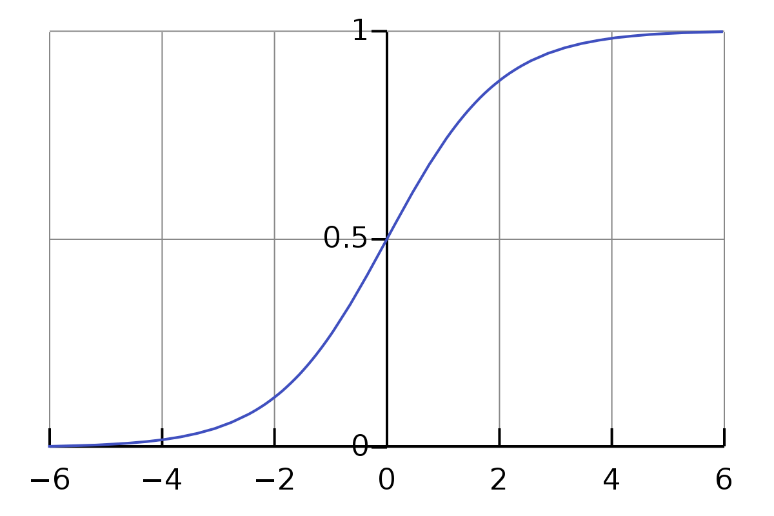


Figure 7 La fonction sigmoid (Saleem, 2023)

Donc pour x1 = 1 et x2 = 0, y = 0.12, donc il y a une erreur puisque la sortie doit être 1.

* Calculer l’erreur

Pour l’erreur rien ne va changer nous allons utiliser la Mean Square Error :

* Backpropagation

Nous voici près pour la rétropropagation, seulement ici nous allons mettre à jour trois poids à savoir w0, w1, w2.

Même si les formules restent les mêmes, ne prenons encore rien pour acquis, ici la valeur de la dérivée partielle va changer étant donné qu’on a introduit une nouvelle fonction, celle d’activation, nous allons de facto nous retrouver avec trois membres dans le calcul de dérivée partielle.

Il suffira de faire les calculs comme nous l’avons fait avec la régression linéaire pour mettre à jour les trois (3) poids. Apres avoir fait tous les calculs, nous allons avoir un tableau comme celui-ci

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X0 | X1 | X2 | W0 | W1 | W2 | Y |  | OU |
| 1 | 0 | 0 | -2.2121 | 5.41528 | 5.41528 | -2.2121 | 0.099 | 0 |
| 1 | 0 | 1 | 3.20318 | 0.961 | 1 |
| 1 | 1 | 0 | 3.20318 | 0.961 | 1 |
| 1 | 1 | 1 | 8.61846 | 0.999 | 1 |

La régression polynomiale

Ci-après un tableau de valeur et sa représentation graphique.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Y | 0.1 | 6.9 | 7.2 | 20 | 28 | 32 | 53 | 62 | 78 | 104.5 |

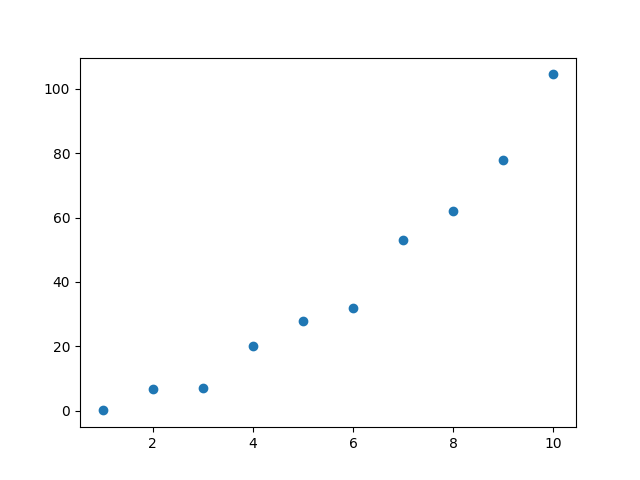


Figure 8 Représentation de points

Apres avoir fait passer ces données dans un modèle de régression linéaire, le résultat obtenu n’était pas à la hauteur de nos attentes.

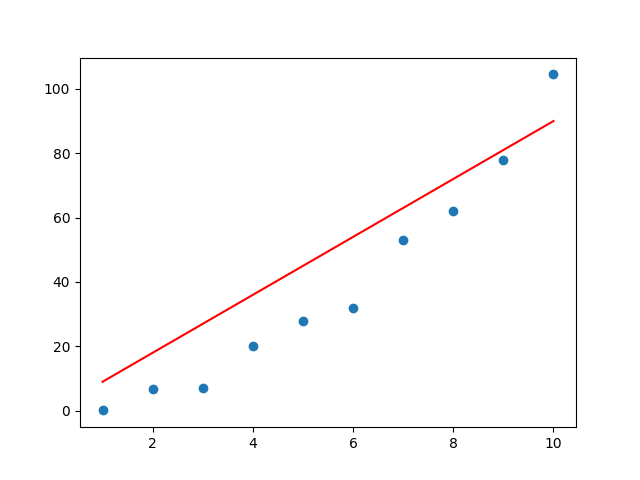


Figure 9 Teste de régression linéaire

Nous voyons que le modèle linéaire ne marche pas sur ces données. Conséquence, nous aurons besoin de quelque chose de plus sophistiquée, et ce quelque chose c’est la régression polynomiale, ce type de régression nous permet de représenter une courbe de donnée qui adapte une forme exponentielle. Les étapes de régression polynomiale restent les mêmes que les autres algorithmes mais ses calculs vont changer.

* Forward-propogation

Pour le Forward-pass de la régression polynomiale, nous allons utiliser, une fonction quadratique, c’est-à-dire qui admet une puissance dans la variable. On va parler de degré de la fonction. Plus le degré est élevé plus la fonction pourra être en mesure d’aller chercher des variations.

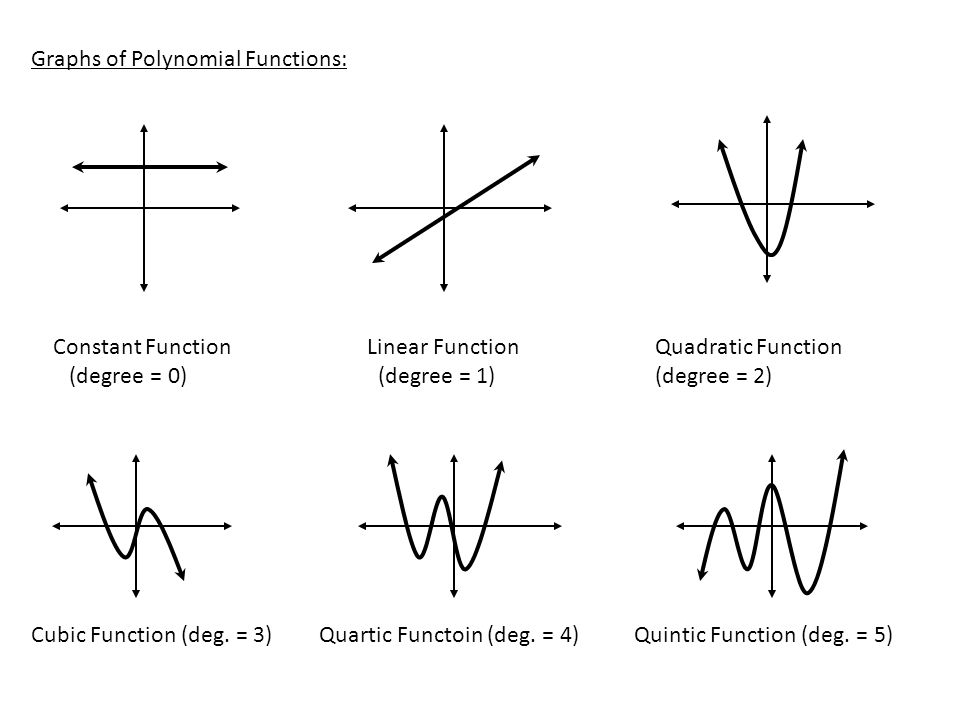


Figure 10 Différents degrés de la régression polynomiale (Graph of polynomial functions, 2020)

Du fait que nous n’avons pas beaucoup de variation dans le tableau, nous allons utiliser le deuxième degré, ainsi notre formule se présente comme suit :

A partir de là, on initialise les poids à 1. C’est le moment de préciser que l’initialisation des poids ne se fait pas forcément avec des uns (1). En vraie, on utilise une fonction de génération de nombres aléatoires, ici nous les initialisons à 1 par souci de simplicité.

Il y a une erreur puisque pour la valeur x = 3, y = 7,2 donc nous allons calculer cette erreur.

* Calculer l’erreur

La fonction d’erreur ne change toujours pas, c’est le MSE.

* Backpropagation

Il est observé une erreur de 17,64, nous allons par la suite retro-propagé cette erreur pour mettre à jour les poids.

La valeur de la dérivée partielle pour w1 se présente comme suit :

Nous avons décidé de prendre w2 car il a la dérivée partielle la plus compliqué à calculer, avec ce calcul établi, nous pouvons passer à l’étape des mises à jour des poids, ne pas oublier de prendre un Learning Rate.

Toutefois, qu’en est-il de notre problème initial, après l’avoir fait passer dans un modèle de régression polynomiale, nous avons trouvé la courbe suivante.

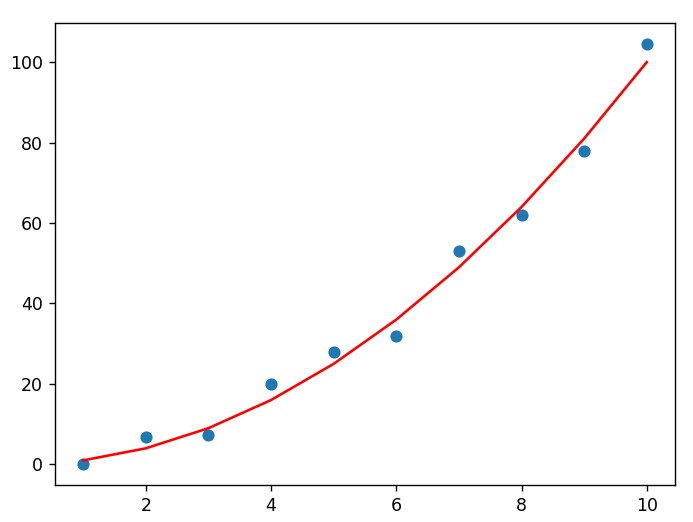


Figure 11 Application d’une régression polynomiale

C’est magnifique, le modèle est parvenu à trouver une corrélation à la presque perfection. Si nous avions plus de variations de la courbe, il nous suffirait d’augmenter le degré et un peu de patience et c’est bon.

Arbre de décision

Pour se faire prenons un exemple concret : ce tableau suivant nous informe si l’individu est sénégalais ou pas à partir de trois (3) attributs, nous allons faire un arbre de décision.

Tableau 4 Exemple de données

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Numéro | Plat | Teint | Taille | Si sénégalais |
| 1 | Riz | Sombre | Grande | Oui |
| 2 | Attiéké | Claire | Petite | Non |
| 3 | Mafé | Sombre | Grande | Non |
| 4 | Riz | Sombre | Grande | Oui |
| 5 | Attiéké | Sombre | Petite | Non |
| 6 | Mafé | Claire | Grande | Oui |
| 7 | Riz | Sombre | Grande | Oui |

* Entropie

L’entropie nous renseigne sur la pureté d’un attribut, si deux classes sont équitablement représentées dans un attribut, on dit que le nœud est impur, conséquence l’entropie est maximale (égale ou proche de 1), si une seule classe est représentée le nœud est pure et l’entropie est minimale (égale ou porche de 0).

Équation Entropie

* Gain d’information (GI)

La première des choses à faire c’est de calculer le gain d’information c’est-à-dire de tous les attributs, quel est celui qui nous renseigne le plus si l’individu est sénégalais ou pas.

Équation 40 Gain d'information

Pour ce qui est de notre exemple :

Calculons l’entropie générale

Gain d’information de l’attribut plat

Si nous répétons les calculs avec les attributs nous allons trouver que

Donc l’attribut plat a le plus grand gain d’information dès lors, il devient l’attribut de plus indicatif et va être à la racine de l’arbre. Notre arbre ressemble à cela après avoir fait tous les calculs :

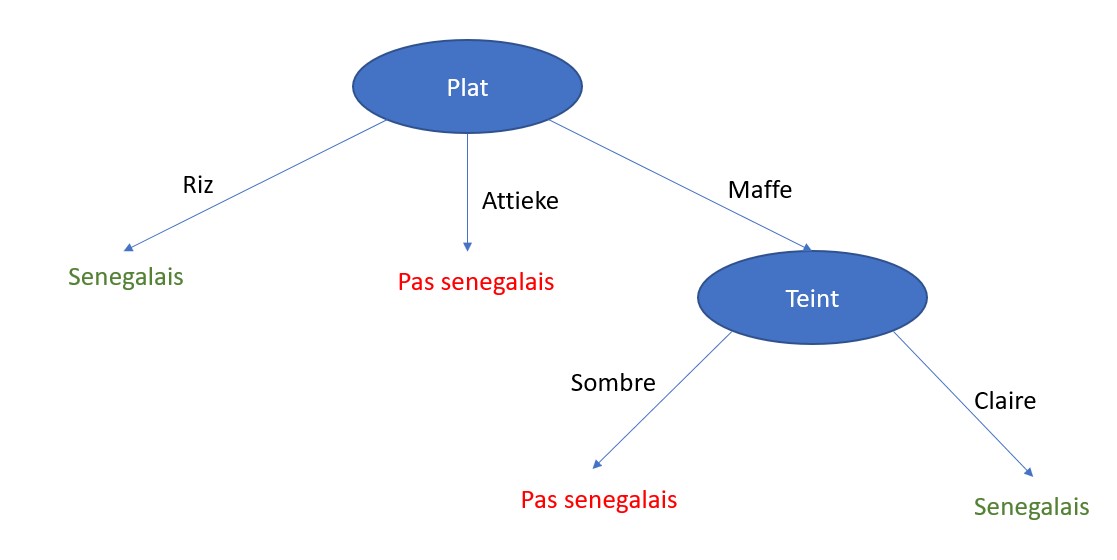


Figure 13 Resultat d'un arbre de decision

Nous constatons qu’il n’est pas nécessaire que tous les attributs soient présents pour prendre une décision ici l’attribut taille n’intervient pas.

SVM

Exemples :

* Si mx + b = -1, y = -1, e = 1 – (-1) (-1) = 0 : pas erreur
* Si mx + b = -1, y = 1, e = 1 – (1) (-1) = 2 : il y a une erreur

Naive bayes

* Exemple : prenons le même exemple du tableau qui essaie de prédire si l’individu est sénégalais ou pas.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Numéro | Plat | Teint | Taille | Si sénégalais |
| 1 | Riz | Sombre | Grande | Oui |
| 2 | Attiéké | Claire | Petite | Non |
| 3 | Mafé | Sombre | Grande | Non |
| 4 | Riz | Sombre | Grande | Oui |
| 5 | Attiéké | Sombre | Petite | Non |
| 6 | Mafé | Claire | Grande | Oui |
| 7 | Riz | Sombre | Grande | Oui |

* Probabilité des valeurs cibles
* Les probabilités des valeurs d’attributs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Plat | Oui | Non |  | Teint | Oui | Non |
| Riz |  |  | Sombre |  |  |
| Attiéké |  |  | Claire |  |  |
| Maffé |  |  |  | | |
|  | | | | | | |
| Taille | Oui | Non |  | | | |
| Grande |  |  |
| Petite |  |  |

* Avec ce tableau nous avons tout ce qu’il nous faut pour classer un nouvel individu. D’ailleurs c’est ce que nous allons faire, classons I1 (Plat = riz, Teint = sombre, Taille = Grande) et I2 (Plat = Attiéké, Teint = claire, Taille = Petite).
* I1 :
* Le modèle nous dit que l’individu est un sénégalais a 100% car la normalisation des probabilités va renvoyer 1 pour sénégalais et 0 pour non sénégalais.
* Si on le fait pour l’individu 2, nous allons trouver (avec les mêmes calculs bien entendu), nous allons trouver que I2 n’est pas du tout sénégalais avec 100% de certitude aussi. Cet exemple est simple mais dans la vie une le modèle va rarement répondre avec une confiance de 100%.

Clustering

Exemple : prenons le tableau suivant :

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
| X | 1 | 1 | 2 | 5 | 5 | 6 | 1 | 1 | 2 |
| Y | 1 | 2 | 1 | 5 | 6 | 5 | 9 | 10 | 9 |

Nous prenons 3 pour la valeur de k, donc nous aurons 3 clusters.

P1, P2, P3 sont choisis et tous les autres vont se classer par rapport au point le plus proche de ses trois. Nous nous retrouvons avec 3 clusters qui se présentent comme suit :

C1 = {P1},

C2 = {P2},

C3 = {P3, P4, P5, P6, P7, P8, P9}.

Calculons les centres de gravité Cg, c’est le point qui représente la moyenne des X et moyenne de Y.

Équation 44 Centre de gravite d'une classe

Cg(C1) = (1, 1), Cg(C2) = (1, 2), Cg(C3) = (3.42, 6.42)

Calculons les distances d pour de tous les points par rapport au centre de gravité, nous allons utiliser la distance euclidienne dans un espace 2D.

Équation 45 Distance euclidienne

Apres calcul, nous avons trouvé le tableau suivant des points et leur distance par rapport au centre de gravité.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
| C1 | 0.0 | 1.0 | 1.0 | 5.66 | 6.40 | 6.40 | 8.0 | 9.0 | 8.06 |
| C2 | 1.0 | 0.0 | 1.41 | 5.0 | 5.66 | 5.83 | 7.0 | 8.0 | 7.07 |
| C3 | 5.94 | 5.04 | 5.60 | 2.12 | 1.63 | 2.94 | 3.54 | 4.21 | 2.94 |

Le constat qui sera fait de ce tableau c’est qu’il y aura du mouvement, et nos clusters deviennent :

C1 = {P1, P2},

C2 = {P3},

C3 = {P4, P5, P6, P7, P8, P9}.

Il ne reste plus qu’à faire la même chose, à savoir recalculer les centres de gravité, recalculer les distances et mettre à jour les clusters. Si nous le faisons assez de fois, nous avons des clusters qui pourrons prédire la classe d’un nouvel individu.

Règles d’associations

* Les données

Une transaction T est considérée comme une itemset, c’est tous les produits qu’un client a acheté en un coup. Nous pouvons avoir quelque chose comme cela.

T1 = {A, B, C},

T2 = {E, F},

T3 = {A, C, E},

T4 = {A, E,}

T5 = {B}.

Nous avons ici cinq (5) transactions avec leurs produits associées. Notre tâche est de trouver le rapport entre l’achat des produits. Si un client achète un produit, dans quelle mesure il achète un autre produit.

* Le support d’un produit

Le support d’un produit, c’est l’occurrence d’un produit dans la transaction sur le nombre de transaction.

Équation 46 Support d'un item

C’est le lieu de parler du support minimum (minsup), c’est le support qu’un produit ou ensemble de produits doit avoir pour rester dans la recherche de règles. Ce minsup est choisi arbitrairement selon le problème posé. Delà, nous avons :

* L’élagage

Nous voyons que le produit F n’a qu’un support de 1/5, donc ne peut pas participer dans une règle intéressante, il va être élagué. Il nous reste {A, B, C, E}.

* La jointure

Cela commence à devenir intéressante, car maintenant nous allons joindre les produits restants et calculer le support des résultats {AB, AC, AE, BC, BE, CE}.

Apres élagage, nous allons nous retrouver avec {AC, AE}.

Une jointure possible et c’est {ACE} qui a un support de 1/5 donc qui va être élagué.

L’ensemble des itemset fréquents est donc {A, B, C, E, AC, AE}, avec cela nous pouvons enfin calculer les règles intéressantes. Une règle est dite intéressante si son support est supérieur ou égal au minsup et sa confiance supérieur ou égale à minconf (la confiance minimum).

Équation 47 Confiance d'une regle

Pour {AC}, calculons-la les règles intéressante avec un minconf de 4/5.

, conf = 2/3, sup = 2/5 : cette règle n’est pas intéressante car son support > minsup et sa confiance < minconf.

, conf = 1, sup = 2/5 : cette règle est intéressante car son support > minsup et sa confiance > minconf.

Conclusion : nous avons pour {AC}, si on achète A il n’est pas forcé que C soit acheté, mais si on achète C, il y a de forte chance que l’on achète A et quand on regarde les transactions, ces conclusions reflètent la réalité. Si je travaille dans cette boutique, je mettrai les produits A à coté des produits C.

ANN

* Architecture du réseau

En pratique, on ne peut pas savoir a priori l’architecture d’un réseau, ce que les scientifiques c’est d’expérimenter jusqu’à trouver le réseau qui le marché de mieux mais consomme le moins d’énergie. Mais nous supposons que ce travail est déjà fait et l’architecture trouvée est :

* + La couche d’entrée avec 3 neurones (il ne faut pas oublier le biais)
  + Une couche de sortie avec trois neurones
  + La couche de sortie avec un neurone
* Feed-forward

Une fois l’architecture définie, on passe à l’essai en initialiser les poids au hasard, nous aurons deux ensembles de poids, ceux qui relient les entrées au cachées, et ceux qui relient les cachées à la sortie.

Équation 48 Calcul de couche cachee ANN

Équation 49 Calcul de la sortie ANN

X : la matrice des entrées

Wxh: la matrice qui relie les entrées et la couche cachée

H : matrice de la couche cachée

Why: la matrice qui relie la couche cachée et les sorties

H : matrice de la couche cachée

Il ne faut pas oublier pour les couches il y a une fonction d’activation, en règle générale, la fonction "relu" est utilisée pour les couches cachées et sigmoid pour les sorties.

Équation La fonction reLu

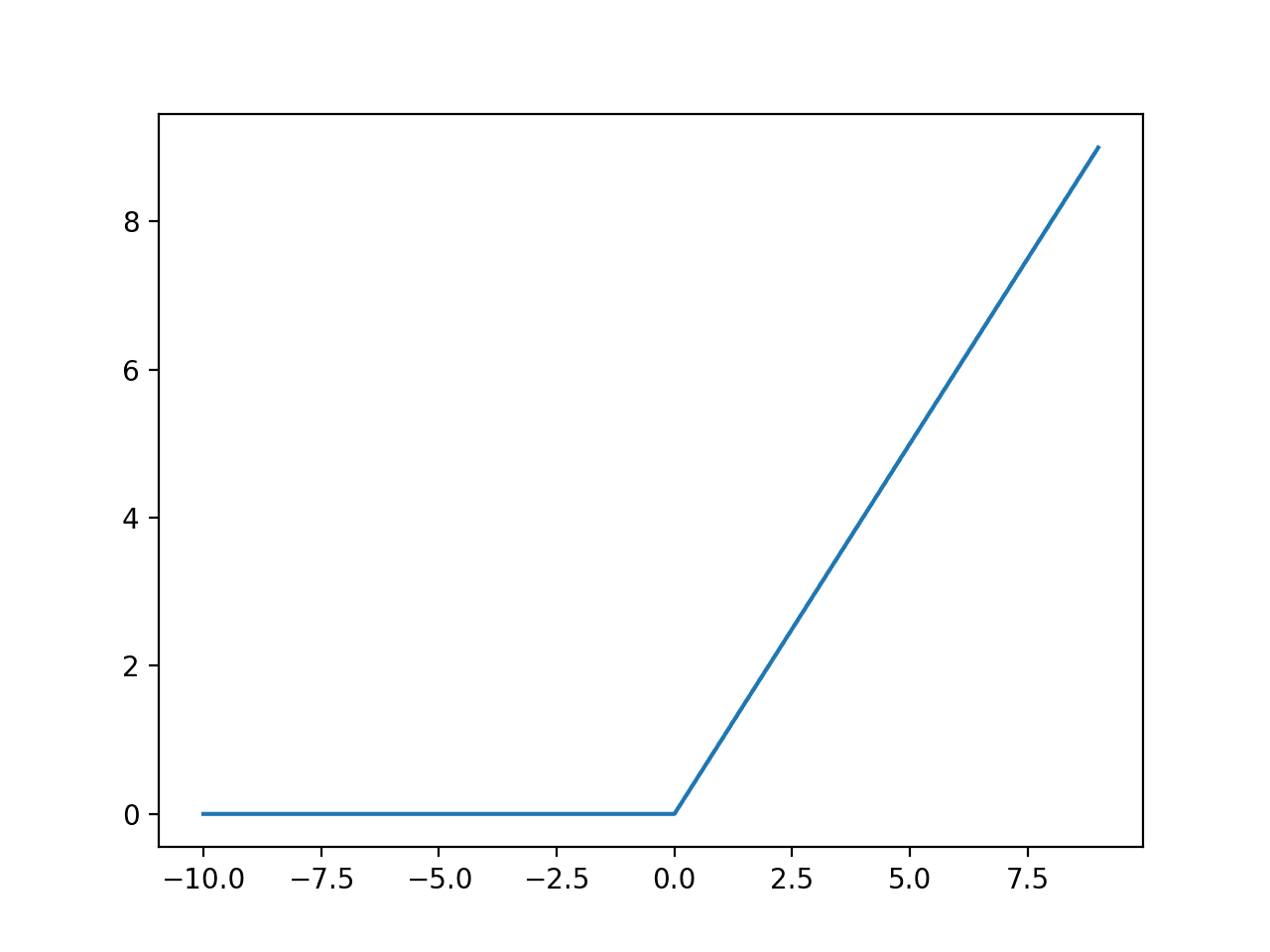


Figure 16 La fonction reLu (Friedman, 2020)

* L’erreur

En Deep Learning il y a plusieurs fonctions d’erreur, il y a le MSE, que nous avons déjà vu, le MAE (mean absolute error), log loss … Mais nous allons encore utiliser le MSE car il fonction très bien pour le problème du XOR. Pour rappel :

* Backpropagation

Une fois que l’on a l’erreur on peut mettre à jour les poids, ici il y aura deux niveaux de mise à jour puisqu’il y deux ensemble de poids qu’il faut mettre à jour. Nous pouvons d’ores et déjà calculer l’erreur pour les neurones cachés.

En résumé, il faut d’abord calculer la matrice relative aux erreurs des couches cachées, ensuite nous mettons à jour les poids des sorties et enfin ceux des couches cachées. C’est comme cela que fonctionnent les réseaux de neurones même dans leurs formes les plus complexes.

CNN

Comment un être humain fait pour reconnaitre les objets qui l’entoure ? nous le faisons tout le temps mais savons-nous comment ce procédé fonctionne dans nos cerveaux. Si nous parvenons à répondre à cette question, il nous sera facile de faire imiter ce procédé par un ordinateur.

|  |  |
| --- | --- |
| Figure Cheikh Anta DIOP | Figure Gaston Berger |

Ici nous avons deux images de personnes célèbres, il s’agit du grand Cheikh Anta Diop et de Gaston Berger. Nous les avons tout de suite reconnus mais comment ? Diop porte des lunettes et Berger pas, Diop et de teint noire et Berger blanc, la moustache de Cheikh Anta Diop et plus touffue que celle de Gaston Berger. Et donc nous avons vu des différences considérables qui nous en permis de les distinguer. Nous allons voir les étapes que l’ordinateur va prendre pour desceller des éléments de différenciation entre ces deux personnes.

* Les filtres

Le docteur Cheikh Anta Diop portait très souvent des lunettes, alors on peut avoir un filtre pour les lunettes, on va chercher sur les images du docteur quelques choses qui ressemble à des lunettes. Ce filtre va être représenté sous forme de matrices et nous allons boucler dans l’image pour le chercher. Si nous le trouvons, on dit qu’il y a des lunettes. Par contre, porter des lunettes de suffit pas pour dire que c’est Dr. Diop, il nous faut trouver d’autres filtres, il va y avoir autant que nécessaire. Mais comment cela se passe en pratique

|  |  |  |
| --- | --- | --- |
| 0.5 | 1 | 2 |
| 6 | 2.3 | 4.1 |
| 2.8 | 0 | 6 |
| 0.14 | 3.51 | 14 |
| 95.2 | 47. | 5 |
| 2 | 15.96 | 4 |

Considérons ce tableau comme les pixels de notre image, nous pouvons avoir un filtre , il nous suffit de faire une multiplication élément par élément, avec chaque matrice qui a les mêmes dimensions que notre filtre à travers tout le tableau. Ensuite, on divise le résultat par le nombre d’éléments dans le filtre. Si vous regardez bien ce filtre est présent dans le tableau, alors nous aurons :

La valeur est plus grande que la valeur maximale du filtre, cela veut dire qu’on a trouvé ce filtre, il faut faire de même pour tous les autres filtres, on obtient ce qu’on appelle un Feature Map.

* L’activation

Une fois ce travail fait, on peut se retrouver avec un Feature map qui ressemble à cela :

|  |  |
| --- | --- |
| -1 | 14.5 |
| 8 | 4.5 |
| -9 | -5 |

Nous allons le faire dans la fonction "relu" et nous allons obtenir cela

* Le réseau de neurone

Il peut y avoir plusieurs Feature map, on va faire ce que l’on appelle le flatten, c’est passer le N dimensions en 1D, maintenant il ne suffit plus qu’à le faire passer ce résultat flatten dans un ANN pour faire la classification.

Pour résumé, il y a deux étapes dans le CNN, le feature extraction et la classification. Le feature extraction c’est tout ce qui est relatif à la recherche d’informations sur l’image.

RNN

Les RNN sont le plus souvent utilisé pour le NLP (Natural Proccesing Language) qui nous permet de comprendre les textes sous forme de nombre. Pour mieux comprendre le NLP, prenons cette phrase : « Il est gentil ». L’ordinateur ne comprend pas le texte et on ne peut pas faire des calculs sur du texte, il va falloir trouver une solution si on veut faire passer cette phrase sur un modèle intelligent. C’est cela le travail du NLP : transformer du texte en un langage compréhensible par la machine, faire les calculs, et le retransformer du langage machine en un langage compréhensible par nous humains.

Exemple : pour la phrase « il est gentil », nous pouvons dire la chose suivante {il : 0, est : 1, gentil : 2}, de ce fait notre phrase devient « 0 1 2 » et il est possible de faire nos calculs. Bon ! dans la vraie vie, les scientifiques utilisent des algorithmes bien plus sophistiqués mais c’est juste pour la compréhension.

Comment les RNN fonctionnent ? Eh bien, presque de la même manière que les ANN, la seule et unique différence c’est la mémoire.

* La conversion en nombre

Avant de commencer le travail, il faut toujours convertir le texte en nombre, pour être plus précis en vecteur. L’exemple donné où l’on remplaçait les mots par des chiffres s’appelle le label encoding, mais y en a d’autres plus utilisé.

* One hot encoding
* Bag of word
* TF-IDF
* Word embedding
* …
* Feed forward

Une fois que les mots ont été convertis, ils peuvent être passés dans le modèle mot par mot mais en donnant la sortie du mot précédent comme input aussi. C’est cela qui nous permet d’avoir une certaine mémoire. Pour l’input N, on lui ajouter la sortie S-1 pour que le modèle se rappelle ce qui était venue avant.

Reinforcement Learning

Les applications du Reinforcement Learning :

* Les voitures autonomes
* La robotique
* Les jeux vidéo
* …
* Annexe 3 : les outils

### C++

Le C++ est un langage de programmation crée en 1985 par l’informaticien danois Bjarne Stroustrup pour pallier aux manquements du langage C qui n’est pas orienté objet. Le C++ est un langage de programmation très utilisé par les développeurs, notamment en ce qui concerne les applications. Il permet d'aborder le développement sous plusieurs paradigmes : programmation générique, procédurale et orientée objet. C'est un langage compilé, ce qui signifie que le code source est traduit en code objet ou binaire pour que la machine puisse l'exécuter. (C++ : présentation du langage de programmation, 2024)

Ce langage de programmation est un langage orienté objet ce qui veut dire il permet de créer des classes. Il est si populaire, on peut donner l’exemple de Google qui l’utilise pour son moteur de recherche, Microsoft qui l’utilise pour World, Excel ou PowerPoint et aussi Autodesk qui l’utilise pour Maya. Pour ce qui est de l’IA, avant l’avènement de Python, les ingénieurs l’utilisaient pour écrire les codes mais son impact est toujours présent. Car derrière presque tous les Framework de Python, qui nous aident dans l’IA, il y a le C++ ou C, le cas de Numpy, Pandas ou Matplotlib.

Les avantages de C++ :

* La performance, rapidité
* La popularité
* La portabilité dans les OS
* L’abondance de bibliothèques
* La programmation orienté objet

Les inconvénients :

* Syntaxe compliquée
* Langage pas du tout pour les débutants

### Python

Python est un langage de programmation créé par Guido Van Rossum. La première version publique du langage est sortie en 1991. Son nom provient de la troupe de comiques anglais les Monty Python.

Python est un langage de programmation dit de “très haut niveau”. Cela signifie qu’il possède un haut niveau d’abstraction par rapport au langage machine. Pour le dire très simplement : plus un langage de programmation est de “haut niveau”, plus sa syntaxe se rapproche de notre langage (l’anglais) plutôt que du langage machine. Un langage de haut niveau est donc plus facile à comprendre et à utiliser qu’un langage de plus bas niveau.

Certains langages (comme Python) utilisent un interpréteur comme traducteur tandis que d’autres utilisent un compilateur.

Un interpréteur se distingue d’un compilateur par le fait que, pour exécuter un programme, les opérations d’analyse et de traductions sont réalisées à chaque exécution du programme (par un interprète) plutôt qu’une fois pour toutes (par un compilateur). (Introduction à Python, 2024)

Les avantages de Python :

* Facile à utiliser
* Sécuriser
* Très populaire
* Compatibilité avec d’autres langages
* Possède beaucoup de bibliothèque pour le Machine Learning

Les limites de Python

* Temps d’interprétation très lent
* Mauvaise présentation des erreurs

### SQL

Structured Query Language (SQL) est un langage de gestion de données sous forme de base de données. Il est utilisé pour gérer des bases de données relationnelles avec ces quatre (4) actions principales que sont le CRUD (CREATE, RETREIVE, UPDATE ET DELETE).

De manière simple SQL va nous permettre de créer des bases de données en utilisant un système de gestion de base de données comme PostgreSQL, Oracle, Maria DB mais nous allons utiliser MySQL. Il va être créer une base de données locale pour stocker les états financiers avec lesquels nous allons travailler.

## Les bibliothèques et Framework

Un Framework est en ensemble de fonction prédéfinie dans un langage de programmation nous permettant de faire une action bien précise. Pour une tache bien définie, si nous avons un Framework, il n’est pas nécessaire de commencer de zéro puisque certaines fonctionnalités sont déjà implémentées. Les Framework ont été créer pour les tâches complexes qui nécessite beaucoup de compétence, ainsi même les développeurs de niveau moyen peuvent créer des programmes avancés, ce qui va servir à la productivité.

Il y a différents Framework pour différents domaines informatiques (développement web, mobile, logiciel…), mais nous allons présenter les Framework qui vont nous aider dans le développement de modèle intelligent et le développement d’interface graphique, ils sont tous liés soit à Python ou à C++.

### Scikt-learn

Scikit-learn est une bibliothèque de Python qui a commencé en 2007 avec le Google Summer of Code Project par David Carpaneau.



Ce Framework s’est spécialisé dans le Machine Learning (supervisé et non supervisé) et nous donne des fonctions pratiques pour le développement de modèles. Scikit-learn supporte parfaitement des domaines comme la classification, la régression, le clustering …

Pour ce qui est de nos modèles, nous allons l’utiliser pour faire la prédiction des éléments des états financiers, la représentation de texte et bien d’autres.

### Spacy

Spacy est un Framework de python qui est utilisée dans le NLP, il nous permet tout simplement de raffiner du texte avant sa modélisation. La raison est que du texte écrit par un humain, avant de faire sa représentation en chiffre, va contenir du bruit. On entend par bruit tous les éléments, mots, vocabulaire qui ne nous sert pas. Exemple dans le langage français, il y a une conjugaison très lourde, un verbe comme « faire » peut donner fait, faite, furent etc. Spacy peut nous aider à trouver la racine de ses mots. Il peut aussi nous aider à tout mettre minuscule, supprimer les ponctuation…



### Gensim

Gensim est un Framework Python avec lequel, il est possible de faire directement de la classification de texte. Dans son fonctionnement, ce Framework utilise les modèle Word2vec et aussi Fasttext. Ces derniers sont des techniques qui sont appelé Self Supervided Learning ou l’on utilise un texte pour générer les inputs et les outputs.



Avec ce Framework on peut calculer la similarité entre deux mots, voir faire des calculs sur des mots, oui des calculs des sur des mots.

### Tensorflow

Si Scikit-Learn est une bibliothèque de Machine Learning, Tensorflow en est une spécialisée sur le Deep Learning. Cette bibliothèque a été développée par Google dans le but de permettre aux experts mais aussi de débutant d’avoir un environnement pour travailler dans le Machine Learning en général.



De tous les Framework que nous avons présentée, Tensorflow est très probablement le plus puissant car nous permettant de faire ce que tous les autres font. C’est un outil tout en un avec ses forces et ses faiblesses.

En plus de tout cela, il nous permet de faire du Computer Vision avec les CNN, du NLP avancée avec la RNN, les LSTM, de créer des API pour le déploiement et bien d’autres.

### Pandas

Pandas est un Framework Python très pratique dans le développement de modèle. Il est utilisé dans le travail a priori, le Feature Engineering. Avec pandas, nous pouvons importer des fichiers CSV, vérifier les données manquantes, les outliers…



Avant chaque développement de modèle, Pandas va certainement intervenir, ce Framework supporte les statistiques qui pourront nous permettre de mettre les données dans un format acceptable par l’ordinateur.

### Numpy

Nativement, les structure de type tableau n’existe pas en Python, il y a des listes en Python pour le remplacer. La différence entre ces deux c’est que les tableaux acceptent un seul type de donnée et taille ne varie pas or, les listes acceptent plusieurs types et sa taille peut varier.

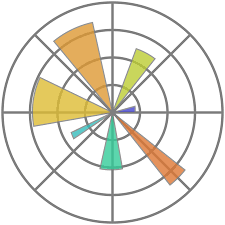


C’est là qu’intervient Numpy pour permettre d’utiliser des tableaux des Python, qui sont bien plus rapide à exécuter. En plus de cela, Numpy a un excellent support de l’algèbre linéaire, les matrices, les vecteurs et autres domaines mathématiques.

Il est nécessaire d’ajouter que Numpy, bien qu’utilisé en Python est écrit en langage C qui est plus puissant et plus rapide que le Python.

### Matplotlib

MatplotLib est un Framework de visualisation avec Python, il sert à tracer des courbes en utilisant Numpy ou Pandas. La visualisation peut intervenir avant et après le modèle, soit pour les comprendre les données brutes, soit pour vérifier les résultats.



La visualisation est en train de devenir une science à part entière, donc Matplotlib est utilisé dans des domaines autres que le Machine Learning, notamment dans le développement d’interface graphique que nous allons voir.

### Qt

Qt est une bibliothèque de C++ cross plateforme lancée en 1995, et qui est complètement gratuit. Qt nous permet de créer des interfaces très avancées et dans un IDE et simple à comprendre et à utiliser. Avec cette bibliothèque, il n’est pas nécessaire de savoir coder pour créer des interfaces graphiques car il y a la possibilité de créer des widgets avec du glisser-déposer.



Pour ce qui nous concerne nous allons bien évidemment l’utiliser pour l’interface qui va accueillir les clients. Mais un logiciel mais pas du web.

Puisque cette application n’a pas pour vocation d’être déployé dans le cloud, le web n’est pas nécessaire. De plus, développer l’interface graphique de cette manière nous donne une certaine sécurité car il n’y aura pas de brèche que des personnes extérieures à l’organisation peuvent utiliser pour accéder aux données sensibles.