
PyNews

Release 0.1

Feb 28, 2020

| | | |
|-----------|----------------------------|-----------|
| 1 | About | 3 |
| 2 | Dependencies | 5 |
| 3 | Instalation | 7 |
| 4 | Set Up | 9 |
| 4.1 | Data Processing | 9 |
| 4.2 | Training | 10 |
| 4.3 | Testing | 11 |
| 5 | Example | 13 |
| 6 | Creation | 21 |
| 6.1 | Data Processing | 21 |
| 6.2 | Feature Tuning | 21 |
| 6.3 | Hyperparameters | 21 |
| 6.4 | Architecture | 22 |
| 7 | Evaluation | 25 |
| 7.1 | Result | 25 |
| 7.2 | Conclusion | 26 |
| 8 | Data module | 27 |
| 9 | Model module | 29 |
| 10 | Trainer module | 31 |
| 11 | Eval module | 33 |
| 12 | Indices | 35 |
| | Python Module Index | 37 |
| | Index | 39 |

A python package for newspaper classification

PyNews is a Natural Language Processing (NLP) short project made at the University of Oslo.

CHAPTER 1

About

This project concerns news article classification using Bag of Words (BoW). A feed forward neural network was trained to predict one of the 20 sources. The **PyNews** package implemented in python was used to explore the different possibilities and training hyper parameters, with the help of PyTorch. We have trained and compared five different architectures on a training and development set. The main focus of this report is to analyse the influence of the number of hidden layers on the model's performance and time efficiency. From these different structures, we selected the best one and trained it three times and will be used to predict unseen data.

CHAPTER 2

Dependencies

PyNews was made in Python 3.7 and uses multiple machine learning libraries :

- PyTorch
- Numpy
- Pandas
- Scikit-Learn
- Pickle
- Time

CHAPTER 3

Instalation

To use this package, clone the repository at <https://github.uio.no/arthur/pynews> on your laptop and from the root folder, run on the commandline :

```
pip install .
```

This will install the package in your python environment and download all the latest dependencies. You can know use and tweak the parameters of pynews' models.

PyNews introduces basic steps for Natural Language Processing (NLP) analysis. First, you may want to clean up the raw data, from *The Signal Media One-Million News Articles* which contains articles from september 2015. This dataset contains Bag of Words (BoW) about 75.000 documents from 20 sources.

4.1 Data Processing

The dataset can be found in the *data* folder, called *signal_20_obligatory1_train.tsv.gz*.

The dataset is made of Part of Speech (POS) tags for every words, meaning that the type (“NOUN”, “VERB”, “ADJ” etc.) are stacked on each words with an underscore. A first step was then to split the words in half, to keep only the word and not its type. Then, the BoW and the vocabulary can be created.

These functionalities are coded in the *pynews.data* module.

```
1 import torch
2 from pynews import NewsDataset
3
4 PATH = "data/signal_20_obligatory1_train.tsv.gz"
5 # Limit of the vocabulary
6 VOCAB_SIZE = 3000
7
8 # Create the PyTorch Dataset object from the dataset
9 dataset = NewsDataset(PATH, vocab_size = VOCAB_SIZE)
```

You can have access to the documents informations from the attributes of the *NewsDataset* class.

```
# The shape of the dataset
>>> dataset.shape
(75141, 3000)

# The classes of the dataset
>>> dataset.classes
```

(continues on next page)

(continued from previous page)

```

['4 Traders', 'App.ViralNewsChart.com', 'BioSpace', 'Bloomberg', 'EIN News',
 'Fat Pitch Financials', 'Financial Content', 'Individual.com',
 'Latest Nigerian News.com', 'Mail Online UK', 'Market Pulse Navigator',
 'Marketplace', 'MyInforms', 'NewsR.in', 'Reuters', 'Town Hall' 'Uncova',
 'Wall Street Business Network', 'Yahoo! Finance', 'Yahoo! News Australia']

# The BoW inputs
>>> dataset.input_features
tensor([[0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        ...,
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.]])

# The gold classes
>>> dataset.gold_classes
tensor([16,  0, 12, ..., 19, 11, 12], dtype=torch.int32)

```

You can now save the vectorizer so your Bag of Words features will always be in the same order.

```
1 dataset.save_vectorizer("vectorizer.pickle")
```

It is recommended to split the dataset in two parts : one used to train the model, the other to evaluate and test it.

```

1 # Define your split ratio
2 SPLIT = 0.9
3
4 train_size = int(SPLIT * dataset.shape[0])
5 dev_size = dataset.shape[0] - train_size
6
7 train_dataset, test_dataset = random_split(dataset, [train_size, dev_size])

```

4.2 Training

Before training the model, divide your dataset in batches and load it with the PyTorch class :

```

1 # Divide your data in batches of size BATCH_SIZE
2 BATCH_SIZE = 32
3
4 train_loader = DataLoader(dataset      = train_dataset,
5                           batch_size = BATCH_SIZE,
6                           shuffle    = True)

```

Then, create your model or use the *NewsModel* one, and define your loss function and optimizer.

```

1 from pynews import NewsModel
2
3 # Define the hyperparameters
4 EPOCHS = 250
5 LEARNING_RATE = 0.09
6 WEIGHT_DECAY = 0.01
7

```

(continues on next page)

(continued from previous page)

```

8 # Create a Feed Forward neural network
9 # with 3 hidden layers
10 # of 150 neurons each
11 num_classes = len(dataset.classes)
12 model = NewsModel(VOCAB_SIZE, 150, 150, 150, num_classes)
13
14 # Loss function
15 criterion = torch.nn.CrossEntropyLoss()
16 Optimizer = torch.optim.SGD(model.parameters(), lr = LEARNING_RATE, weight_decay =
↳WEIGHT_DECAY)

```

You can now train the model with :

```

1 from pynews import Trainer
2
3 # Create your trainer for your model
4 trainer = Trainer(model, train_loader)
5
6 # Run it with the hyper parameters you defined
7 train_losses = trainer.run(criterion, optimizer, EPOCHS, LEARNING_RATE)

```

4.3 Testing

Now that your model is trained, evaluate it on the test dataset.

```

1 # Load the dataset
2 train_loader = DataLoader(dataset = test_dataset,
3                           batch_size = BATCH_SIZE,
4                           shuffle = True)
5
6 # Evaluate the model
7 test_accuracy, test_predictions, test_labels, confusion_matrix = eval_func(train_
↳loader, model)
8 # Get the per class accuracy
9 per_class_accuracy = confusion_matrix.diag() / confusion_matrix.sum(1)
10 # Compute the precision, recall and macro-f1 scores
11 precision = precision_score(test_labels, test_predictions, average='macro')
12 recall = recall_score(test_labels, test_predictions, average='macro')
13 macro_f1 = f1_score(test_labels, test_predictions, average='macro')

```

```

# Global accuracy
>>> test_accuracy
0.5281437125748503

# Per class accuracy
>>> per_class_accuracy
tensor([[0.4314, 0.8500, 0.1333, 0.2852, 0.8547, 0.2279, 0.1297, 0.5329, 0.5388,
         0.5556, 0.1435, 0.2082, 0.3446, 0.7043, 0.5000, 0.1399, 0.4604, 0.1401,
         0.3069, 0.4850]])

# Precision, recall and macro-F1 scores
>>> precision
0.4126984126984127
>>> recall

```

(continues on next page)

Example

In this example, we will try to predict the source of an unknown document.

Open your model and the *vectorizer.pickle* to process the data exactly the same way as the training data.

```

1 import numpy as np
2
3 import torch
4 from torch.utils import data
5
6 # Load your pytorch model
7 model = torch.load("your_model_path.pt")
8 # Open your vectorizer file
9 with open("vectorizer.pickle", 'rb') as f:
10     text_vectorizer = pickle.load(f) # Loading the vectorizer

```

Download the article from one of the 20 classes used to train the model. For example, we will try to predict the source of this article (<https://myinform.com/banking-real-estate/best-bank-in-canada/>).

```

1 document = """
2     Best Bank in Canada
3
4     There was a time when you had to physically haul yourself to the closest
5     ↳branch that your bank was operating out of in the area just so that you could
6     ↳withdraw some money or check the balance remaining in your bank account. With the
7     ↳advent of technology and the widespread percolation of the internet, the banking
8     ↳sector has seen some major changes, which brought about a complete shift in how you
9     ↳store your money and the ways in which you gain access to it. Almost all banks --
10    ↳both large and small -- have their own website and web portals where transactions
11    ↳can be carried out by the customers who sign up for these services. This not only
12    ↳makes your life easier, but also gives you a secure way of keeping an eye on your
13    ↳funds.
14
15    The main problem that many people face in this context is that there are
16    ↳way too many options available for them to choose from, which can be an
17    ↳overwhelming task to sort through. If you are facing such issues and want some help
18    ↳in finding out what is the Best Bank in Canada for you to choose from, you have
19    ↳found yourself at the right place. During the course of this guide, we (continues on next page)
20    ↳looking at what online banks really are, why we need them and some of the best
21    ↳options that are available for you -- so keep reading!

```

(continued from previous page)

6
7 What Are Online Banks?

8 'Online banks' is a term that is used quite loosely to cover many of the
↪ financial institutions that allow you to carry out your banking through internet-
↪ based platforms. For example, most regular banks have a dedicated online section,
↪ where registered users can utilize an electronic payment system to conduct an
↪ impressive range of financial transactions, which essentially work in the same way
↪ as regular banks. Similarly, there are also a bunch of online only banks that are
↪ popping up around the world. While these banks do not have actual physical branches,
↪ that you can visit to get your work done, you can conduct all your business from
↪ the comfort of your home or office by securely logging on to the company's platform.

9 While many people were apprehensive about the security aspect of online
↪ banks when they were first introduced into the market, such financial institutions
↪ have become the base on which our financial transactions are based now. If you are
↪ still wondering whether or not online banks are the right way forward for you, the
↪ next section will give you a better insight into what are the main advantages
↪ associated with using online banks for your transactions.

10
11 Why Should I Use an Online Bank?

12 While most people prefer to keep an account in a regular bank at all
↪ times, many are starting to experiment by opening accounts in these online banks
↪ that minimize face-to-face interactions with bank staff. If you have ever tried
↪ online banking, you already know how easy it is to get all your work done sitting
↪ in one place when you would otherwise be moving from one counter to the other. The
↪ case for online banking is based on a three-pronged approach to its benefits, all
↪ three of which will be discussed below in greater detail:

13 Convenience -- First and foremost, online banks give you access to your
↪ account from anywhere in the world. You can use your computer, smartphone or tablet
↪ to carry out whatever business you have without having to search for ATMs to
↪ withdraw money from or physical bank branches to seek out a staff member. How many
↪ times has an important task come to your mind only to be pushed back because
↪ banking hours were over? That will never be the case with online banking, as you
↪ have access to your account round the clock.

14 Pricing -- There are a few conventional banks that will offer you "low-
↪ fee" accounts, but in most cases, these accounts will be accompanied by some
↪ monthly fees and charges that can really wreak havoc on your pocket over time. When
↪ you choose to go for online banking, you don't just gain access to low-fee options,
↪ but also a bunch of different no-fee accounts, the best of which will be discussed
↪ in a later section. The rationale behind this is that why should you be paying to
↪ gain access to the money you have saved yourself!

15 Additional Features -- Conventional banking doesn't give you the freedom
↪ to deposit cheques through your phone or simply send out an email when you want to
↪ send money. With the help of features like mobile deposits and Interac e-transfers,
↪ online banks have an edge over the regular banks -- and these aren't the only
↪ services you would be able to gain access to!

16
17 The Best Bank in Canada

18 There are a bunch of different things that you should be looking at if
↪ you want to choose the best online bank for your needs out of the large number that
↪ are out there right now. While the pricing offered by one bank may seem very
↪ impressive, another one may be able to give you access to special features that
↪ suit your individual needs in a better manner. Regardless of what the case may be,
↪ here is a list of the best online banks in Canada, along with an emphasis on what
↪ they do best so that you can choose one that is more in line with your needs:

19
20 Tangerine

21 The first name on the list of the best online banks in Canada is the
↪ online-only vertical of the Scotiabank called Tangerine. When it comes
↪ out regular transactions like routine withdrawals, transfers and payment of bills
↪ and more, this bank will give you deals and facilities that are way above the
↪ industry standard -- all for free! While the company does offer a decent
↪ interest rate for savings accounts, you can avail the promotional offer that allows
↪ you to earn 2.75% interest on your first savings account with Tangerine for a
↪ period of 6 months.

(continues on next page)

(continued from previous page)

22 The company also offers a great no-fee chequing account, which will be
↳discussed in greater detail in the next section, as well as a Tangerine Money-Back
↳Credit Card on which you can earn 4% Money-Back Rewards with certain conditions.
↳The best part about this online bank is that there are no annual fees involved when
↳you apply for this chequing account or credit card, which is often an issue with
↳conventional banks and platforms. Finally, Tangerine will go out of its way to help
↳you with any issues with the help of its award-winning customer service that can be
↳availed online, on the phone or through social media.

23 Scotia OnLine

24 While Tangerine is a great online bank for you to turn to, there is
25 ↳another good option for you if you want a company with more experience. Tangerine's
↳parent company Scotiabank, one of the "Big 5" banks in the country, has an online
↳interface that allows you to make the most of the bank's varied features and
↳services. If you are a person who chooses his or her banking institutions in terms
↳of the digital interface that they offer, it doesn't really get better than Scotia
↳OnLine.

26 One thing that you need to keep in mind is that you will be expected to
↳pay a monthly service fee, which is the norm for some of the larger conventional
↳banks. This fee can vary from anywhere between \$3.95 and \$10.95, but if your
↳account has a minimum balance, or if you are a senior citizen or a student, you can
↳get a discount on this monthly fee. The bank is always offering its clients some or
↳the other special promotion in terms of rewards, so even if you aren't looking
↳actively, you must keep an eye out for these special offers.

27 EQ Bank

28 As more and more banks and financial institutions switch over from
29 ↳conventional branches and ATMs to online platforms, EQ Bank (a subsidiary of
↳Equitable Bank, one of the larger banking institutions in Canada) has developed a
↳model where you carry out your banking and transactions only through the internet
↳using their web portal and smartphone app. This means that you don't need to get
↳your hands on a cheque book or a debit card, and can carry out all transactions by
↳moving money between your EQ Bank account and any linked accounts.

30 If you are ready to get rid of the conventional shackles of banking, you
↳can give EQ Bank a shot because of some of the self-evident benefits that are
↳associated with setting up an account with the company. For example, you will not
↳have to pay any monthly fees or charges when you set up any account with the
↳company. In addition to that, the no-fee EQ Bank Savings Plus Account gives you a 2.
↳30% interest, which is fairly impressive for the industry. The only downside of
↳opting for an account with EQ Bank is that you won't get the option of opening a
↳chequing account, which could be a deal breaker for some.

31 RBC Online

32 Another option for those of you who are looking to open an online account
33 ↳with one of Canada's "Big 5" banks is RBC Online, the online platform offered by
↳Royal Bank of Canada. RBC Online gives you a wide variety of accounts to choose
↳from, all at different levels and varying fee structures. On average, you can
↳expect yourself to pay anywhere between \$4 and \$30. This is a great option for
↳those of you who want to bundle up a bunch of different RBC products like
↳investments, mortgage and credit cards and cut down on service charges. However,
↳you will not be able to get any fee waivers by maintaining a minimum account
↳balance, which is otherwise an option with many other similar banks.

34 All in all, RBC Online is a great option for people who are still pulled
↳towards the sense of security that is associated with traditional banking
↳institutions, but also want to get things done online, for example, paying bills,
↳transferring funds and so on. That being said, you cannot expect the bank to give
↳you any other features or services that make it stand out from the other options.

(continues on next page)

(continued from previous page)

BMO Online

If you are a student, senior or a serving or retired military personnel, you must look at the services offered by another "Big 5" bank in Canada -- the Bank of Montreal's BMO Online service. As is the case with the banks in this category -- some of which have been discussed earlier in this section -- BMO Online will get you similar standardized features like interest rates that match the industry norm and so on. If you have a Savings Builder Account, for example, you will be able to get anywhere up to 1.6% interest. The monthly fees charged for the different accounts and services is also pretty standard -- between \$4 and \$30 -- but you can avail of a wide range of special discounts that are only made available to students, seniors and armed forces personnel.

While this online platform isn't really known for its innovative approach to online banking, it will give you the freedom to handle all your financial moves from a well designed web portal and smartphone app, which is essential for online banks.

TD Online

If you are looking to open an account with an online bank, you probably already know how important it is to have a well-designed mobile app that lets you take all your financial decisions, transfer of funds as well as bill payments while you are on the go. If a great mobile app is your requirement from an online banking service, Canada's Toronto-Dominion Bank has an online platform called TD Online, which is virtually unmatched in terms of mobile app performance.

First and foremost, it is important to note that TD Canada Trust is not going to give you anything exceptional when it comes to different account options or related interest rates. However, it does give you the option of setting up a good enough no-fee savings account. The mobile app, which is the real winner with TD Online, is filled with a bunch of attractive features that sets it apart from the mobile apps put in place by the competitors. For example, the app comes with features like TD MySpend, which is a tool designed to ensure that you have a platform to track your spending habits and make improvements to your lifestyle.

CIBC Online

The last member of the "Big 5" banks to make it to the list of online banks is the Canadian Imperial Bank of Commerce and its CIBC Online platform. The best part about this bank and its platform is that it has been designed to look into the needs of different people who are all at different stages of their lives. This means that if you are a senior looking to open an account with one of the bigger banks in the country, you can look up CIBC's five different account types that have been specially tailored to the needs of seniors. These accounts can be regular chequing and savings accounts or full-fledged bundles that get you access to elite-level credit cards or just simply a US dollar account.

More and more seniors feel themselves gravitating towards CIBC because of the sense of security that a big bank offers to people who weren't born in the internet generation. While there may be tons of newer options on the market, sticking to a well-known name may help them feel more secure.

FirstOntario

In this section, we will be deviating from the regular banking institutions to include a credit union -- the only one on this part of the list -- that gives you options to set up chequing accounts as well as savings accounts that may be free or with small associated charges or fees. With the help of the FirstOntario, you can carry out all the tasks you would traditionally be able to conduct with the others on this list. For example, you can use the banking option with FirstOntario to carry out bank transfers, pay your bills online, withdraw money, conduct Interac e-Transfers and even give out personal cheques.

(continues on next page)

(continued from previous page)

In addition to all these basics, FirstOntario also gives you access to a bunch of different online tools that set it apart from the others in the field. For example, you can use the FirstOntario features to round up when you make a purchase, automatically saving the balance that remains. This acts like a very helpful saving tool.

Simplii Financial

Another online banking option that is affiliated with CIBC is Simplii Financial. This is a bank that gives you the option to conduct all your business without ever having to visit a physical branch because there aren't any branches at all! This isn't a platform that gives you a ton of different options when it comes to account types. There are only two types of accounts that you can open -- the chequing account and the savings account. The chequing account will be discussed in greater detail in the next section, but both these accounts can be opened without having to pay any monthly fees or charges!

You can do everything from paying bills to transferring funds and making cheque deposits (entirely paperless) with the help of the web and mobile platforms that Simplii Financial offers to its consumers. You can learn more about this in the next section.

motusbank

The last name on this list is that of the digital bank set up by another credit union -- motusbank by Meridian. Launched not too long ago, this online bank has created waves across Canada because of the wide range of products that it has to offer to its consumers. The positives don't just end here, as the company also offers interest rates that can help you make better financial decisions. With no annual or monthly fees associated with setting up an account as well as no minimum balance requirement, motusbank is a great option not just for chequing accounts but also if you need a mortgage. The latter is also the main reason why motusbank has made it to this list of the best online banks in Canada. If you are looking for personal loans, a mortgage or opening an account, this is one option that you must consider.

Best No-Fee Chequing Accounts

After having looked at the best online banks in Canada, it is time for us to look at the best individual options in front of you when it comes to no-fee chequing account. As opposed to the low-fee options that many regular banks offer to their consumers, these no-fee chequing accounts come without the fear of racking up monthly or yearly fees and charges. For this reason, they are one of the more sought out options for people who want to save some money on such unnecessary expenditure. In the list below, we will briefly touch upon the best no-fee chequing account options for you.

Tangerine No-Fee Chequing Account

You may be familiar with the bank that was known as ING Direct. Revamped under the brand name of Tangerine, this is an online only bank that is owned by Scotiabank and was discussed in greater detail in the previous section. Here, we are going to look deeper into the free chequing account that the company offers to its clients, in addition to the high-interest savings account, which is offered at a 2.75% promotional rate. While we will not be going into the details of the high-interest savings account, you should keep reading to find out more about the chequing account.

First and foremost, the chequing account is offered to you without any monthly fees. This means that you can carry out an unlimited number of transactions with this account, without having to worry about a bunch of additional charges. The company also gives you access to 3,500 of Scotiabank's ATM located across Canada and a total of 44,000 locations across the world. You can handle all your transactions through this well-designed app that the company offers and 24/7 phone support in case of any issue. Tangerine also offers interest payments of 0.15% to 0.65% on the account balance, in addition to the free first cheque book that you are entitled to. Keep in mind that you will be charged an inactivity fee of \$10 if your account doesn't have any activity for 12 months.

(continues on next page)

(continued from previous page)

65 motusbank No-Fee Chequing Account

66 Another great option for those of you who are seeking out an online bank
67 ↳that will give you a no-fee chequing account is motusbank. Owned by Meridian Credit
↳Union, this is one of the newer entries into the category of online-only banks and
↳gives you a bunch of different options to choose from, as was discussed in the
↳earlier section. While motusbank will give you the option to set up a high-interest
↳savings account, as well as mortgages, personal loans and investments, this section
↳will only focus on the no-fee chequing account that motusbank has to offer to its
↳patrons.

68 For no monthly account fee and an impressive 0.50% interest payment on
↳your account balance, this is one of the market leaders when it comes to no-fee
↳accounts by online banks. The company will give you unlimited Interac e-Transfers
↳for free, along with other transactions like debit purchases, withdrawals and bill
↳payments. While this bank also gives you the first cheque book for free, it has
↳only half the number of cheques that Tangerine offers. You also gain access to over
↳3,700 ATMs all across the country, but keep in mind that there will be an
↳inactivity fee of \$30 dollars if the account isn't used for 12 months.

69 Simplii Financial No-Fee Chequing Account

70 The next option on the list is the no-fee chequing account offered by the
71 ↳company that was known as PC Financial in the past. Owned by CIBC, Simplii
↳Financial gives you the option of setting up no-fee chequing accounts, high-
↳interest savings, mortgages and so much more so that you can benefit greatly
↳without the fear of additional charges and fees. The no-fee chequing account, which
↳is the main service that will be discussed in this section, comes with no monthly
↳fees -- as the name suggests -- as well as no minimum balance requirement, which
↳can otherwise act as a hurdle in your path.

72 If you sign up for the no-fee chequing account by Simplii Financial, you
↳will be able to make the most of the free unlimited transactions that the company
↳offers to its customers. These transactions include everything from debit purchases
↳to bill payments that have been pre-authorized, which solves the problem of
↳recurring bills. Along with access to the thousands of CIBC ATMs across the country,
↳you also get as many free Interac e-Transfers as you want to carry out. A great
↳feature offered by company is the free cheque books and the versatile app that lets
↳you control the account seamlessly. An inactivity fee of \$20 will be charged if you
↳don't carry out any transaction for 2 years.

73 Motive Financial

74 Canadian Western Bank's Motive Financial -- known as Canadian Direct
75 ↳Financial in the past -- also offers two different types of free chequing accounts
↳to its consumer base. The first of these is the Motive Chequing account, while the
↳other is the Motive Cha-Ching Chequing account. In addition to this, the company
↳also gives you the option to enroll for savings accounts, investments and a bunch
↳of other services that can make your life so much easier.

76 Moving to the Motive Chequing account, which is going to be our main
↳point of discussion under Motive Financial's offered services, let's look at some
↳of the distinguishing features. First, there are no monthly or yearly fees
↳associated with setting up this account that allows you an unlimited number of
↳transactions. Not only do you get 0.25% interest payment on your account balance,
↳you also gain access -- for free -- to one of the largest ATM networks in the
↳country. You can control your account through an effective and easy to use mobile
↳application and also get your hands on a free cheque book with 50 cheques. The main
↳issue here is that you will be charged a \$1 fee for every Interac e-Transfer and an
↳inactivity fee of \$20 when you make no transactions for 2 years.

77 Manulife Advantage Account

(continues on next page)

(continued from previous page)

```

79         The final option on this list before we conclude is the Manulife
↳ Advantage Account, which is quite different from the others on the list, as it
↳ gives you access to features of both a chequing account as well as a savings
↳ account. While you will be required to maintain a minimum account balance of $1,000,
↳ you can carry out an unlimited amount of transactions like bill payments and
↳ Interac e-Transfers, without any fees or charges. In addition to this, you will be
↳ able to avail high interest rate payments on your account balance. With access to
↳ thousands of ATMs across the country and a smartphone app that allows you to carry
↳ out e-Transfers and deposit cheques, you will not be lacking any of the premium
↳ features offered by other online banks. As is the case with the others, an
↳ inactivity charge -- of $20 -- will be applicable after you do not carry out any
↳ transactions for two years.
80         As the world moves towards greater technological advancements, there is
↳ no aspect of your life that is not touched by the internet -- including the banking
↳ sector. If you are looking to open an account with a good online bank that isn't
↳ just going to give you a no-fee option but also great customer service combined
↳ with a bunch of different features, you can look at the options mentioned above.
↳ All in all, the decision will have to be based on your individual needs, as people
↳ differ in terms of what they need from an online bank. Regardless, you can be
↳ assured that any of the banks mentioned above will strive towards giving you a
↳ fantastic experience!
81         """

```

Extract the Bag of Words features and the vocabulary of the document with the vectorizer.

```

1  # The Bag of Word
2  input_features = text_vectorizer.transform([document]).toarray().astype(np.float32)
3  # Converting the numpy array to pytorch tensors
4  torch_input_features = torch.from_numpy(input_features)

```

At this point, you can now predict the source.

```

1  # Run the model on the unknown document
2  prediction = model(torch_input_features)
3  predicted = prediction.argmax(1)
4  # Classes generated by the training processing
5  classes = ['4 Traders', 'App.ViralNewsChart.com', 'BioSpace', 'Bloomberg', 'EIN News',
6            'Fat Pitch Financials', 'Financial Content', 'Individual.com',
7            'Latest Nigerian News.com', 'Mail Online UK', 'Market Pulse Navigator',
8            'Marketplace', 'MyInforms', 'NewsR.in', 'Reuters', 'Town Hall' 'Uncova',
9            'Wall Street Business Network', 'Yahoo! Finance', 'Yahoo! News Australia']
10 # Get the name of the predicted class
11 predicted_class = classes[predicted]

```

```

>>> predicted_class
MyInforms

```


6.1 Data Processing

The dataset can be found in the *data* folder, called *signal_20_obligatory1_train.tsv.gz*.

The dataset is made of Part of Speech (POS) tags for every words, meaning that the type (“NOUN”, “VERB”, “ADJ” etc.) are stacked on each words with an underscore. A first step was then to split the words in half, to keep only the word and not its type. Then, the BoW and the vocabulary can be created.

These functionalities are coded in the *pynews.data* module.

6.2 Feature Tuning

Before changing the structure of the model, we explored different Bag of Words features implementation by varying the vocabulary size, the preprocessing and building the vocabulary before and after the split. For example, with a vocabulary size of 4000 we did not see improvement in the performance. In addition, using the Part of Speech (POS) tags did not help to optimize the results. Because of difficulties and time restriction we did not create the vocabulary after splitting into training and development parts.

6.3 Hyperparameters

A brief training session to evaluate the performance with different hyper parameters was firstly performed. The hyper parameters used are described on the table below.

Table 1: Hyperparameters

| Parameter | Value |
|-----------------|-------|
| Split Train/Dev | .9 |
| Vocabulary | 3000 |
| Batch size | 32 |
| Learning Rate | .09 |
| Epochs | 250 |

6.4 Architecture

Then, five different models were trained on Saga’s server with different layout. These models differ in their number of hidden layers, and their architectures are presented in the tables below.

Table 2: First Model

| Layers | Neurons | Activation |
|--------|---------|------------|
| Input | 3000 | ReLU |
| Hidden | 150 | Linear |
| Output | 20 | Softmax |

Table 3: Second Model

| Layers | Neurons | Activation |
|----------|---------|------------|
| Input | 3000 | ReLU |
| Hidden 1 | 150 | ReLU |
| Hidden 2 | 150 | Linear |
| Output | 20 | Softmax |

Table 4: Third Model

| Layers | Neurons | Activation |
|----------|---------|------------|
| Input | 3000 | ReLU |
| Hidden 1 | 150 | ReLU |
| Hidden 2 | 150 | ReLU |
| Hidden 3 | 150 | Linear |
| Output | 20 | Softmax |

Table 5: Fourth Model

| Layers | Neurons | Activation |
|----------|---------|------------|
| Input | 3000 | ReLU |
| Hidden 1 | 150 | ReLU |
| Hidden 2 | 150 | ReLU |
| Hidden 3 | 150 | ReLU |
| Hidden 4 | 150 | Linear |
| Output | 20 | Softmax |

Table 6: Fifth Model

| Layers | Neurons | Activation |
|----------|---------|------------|
| Input | 3000 | ReLU |
| Hidden 1 | 150 | ReLU |
| Hidden 2 | 150 | ReLU |
| Hidden 3 | 150 | ReLU |
| Hidden 4 | 150 | ReLU |
| Hidden 5 | 150 | Linear |
| Output | 20 | Softmax |

After we trained all models, there was no specific model that stood out. However, we tried to differentiate them regarding four indicators : the accuracy, macro-F1, precision and recall.

7.1 Result

As shown in the table below, the model 4 presents the best accuracy. Nevertheless, this might not be the only criterion to consider, especially because of the different sources frequency. The Macro-F1 score might be less sensitive to imbalanced class frequencies. In that case, the model 3 performs better even if its precision is not optimal. Due to these performance, we choose the model 3 to push further the training.

Table 1: Models performance

| Model | Accuracy | Macro-F1 | Precision | Recall | Run Time |
|---------|----------|----------|-----------|--------|----------|
| Model 1 | 53.33 | 38.32 | 39.04 | 42.11 | 00:22:27 |
| Model 2 | 52.35 | 32.93 | 36.84 | 31.58 | 00:25:43 |
| Model 3 | 52.81 | 43.59 | 41.27 | 49.44 | 00:27:04 |
| Model 4 | 53.55 | 42.22 | 44.11 | 31.58 | 00:28:40 |
| Model 5 | 53.52 | 42.00 | 46.80 | 43.27 | 00:25:41 |

The mean and standard deviations of the metrics when running the chosen model three times is displayed in the table below.

Table 2: Model 3 performance

| Metric | Average | STD |
|-----------|---------|-------|
| Accuracy | 53.31 | 0.29 |
| Macro-F1 | 34.07 | 4.07 |
| Precision | 33.85 | 4.13 |
| Recal | 37.27 | 5.47 |
| Run Time | 42.73 | 10.28 |

7.2 Conclusion

As shown in the results, the performance are really different from one training to the other. Most likely this is due to us no setting a random seed, leading to different weights, bias and also affecting the optimizer.

class `pynews.data.NewsDataset` (*path*, *vocab_size=3000*)

`NewsDataset` class of different newspaper. The class inherits from `torch.utils.data.DataSet` class.

save_vectorizer (*filename='vectorizer.pickle'*)

Save the `text_vectorizer` method used to transform the dataset into Bag of Words tensors. This file is saved with a binary pickle format.

Parameters **filename** (*str*, *optional*) – Name of the text vectorizer file. The default is “vectorizer.pickle”.

Returns

Return type `None`.

tokenizer (*sample*)

Tokenize a text and extract only the words.

Parameters **sample** (*list*) – Text sample to tokenize.

Returns **sample** – Tokenized text, containing only words.

Return type `list`

class `pynews.model.NewsModel` (**layers_size*)
Newspaper Model.

forward (*inputs*)

Predict the outputs from the given inputs. The ReLU function is used from the inputs to the last hidden layer, then a linear function from the last hidden layer to the outputs.

Parameters `inputs` (*torch tensor*) – Inputs tensor.

Returns `outputs` – Predicted tensor.

Return type torch tensor

class `pynews.trainer.Trainer` (*model, train_loader, model_name='model'*)

Train a model with a dataset.

run (*criterion, optimizer, epochs, lr*)

Method to run the trainer. Use this function to train your model with your data.

Parameters

- **criterion** (*torch.nn.modules.loss*) – The loss function.
- **optimizer** (*torch.optim*) – Method to optimize the model.
- **epochs** (*int*) – Number of iterations.
- **lr** (*float*) – Learning rate.

Returns `train_losses` – List of size epochs, containing

Return type list

`pynews.eval.analyze_confusion_matrix` (*confusion_matrix*)

Analyse a confusion matrix by printing the True Positive (TP), False Positive (FP), ... and the specificity and sensitivity as well for each classes.

Parameters `confusion_matrix` (*torch tensor of size (number_of_classes, number_of_classes)*) – The confusion matrix computed on the test data.

Returns

Return type None.

`pynews.eval.eval_func` (*batched_data, model*)

Evaluate the model on the test data.

Parameters

- **batched_data** (*torch DataLoader*) – The loaded test data.
- **model** (*torch Model*) – The PyTorch model to evaluate.

Returns

- **accuracy** (*float*) – Global accuracy of the model.
- **predicted** (*torch tensor*) – Predicted output.
- **gold_label** (*torch tensor*) – Truth output.
- **confusion_matrix** (*torch tensor*) – Confusion matrix.

CHAPTER 12

Indices

- genindex
- modindex
- search

p

`pynews.data`, 27
`pynews.eval`, 33
`pynews.model`, 29
`pynews.trainer`, 31

A

`analyze_confusion_matrix()` (*in module `pynews.eval`*), 33

E

`eval_func()` (*in module `pynews.eval`*), 33

F

`forward()` (*`pynews.model.NewsModel` method*), 29

N

`NewsDataset` (*class in `pynews.data`*), 27

`NewsModel` (*class in `pynews.model`*), 29

P

`pynews.data` (*module*), 27

`pynews.eval` (*module*), 33

`pynews.model` (*module*), 29

`pynews.trainer` (*module*), 31

R

`run()` (*`pynews.trainer.Trainer` method*), 31

S

`save_vectorizer()` (*`pynews.data.NewsDataset` method*), 27

T

`tokenizer()` (*`pynews.data.NewsDataset` method*), 27

`Trainer` (*class in `pynews.trainer`*), 31