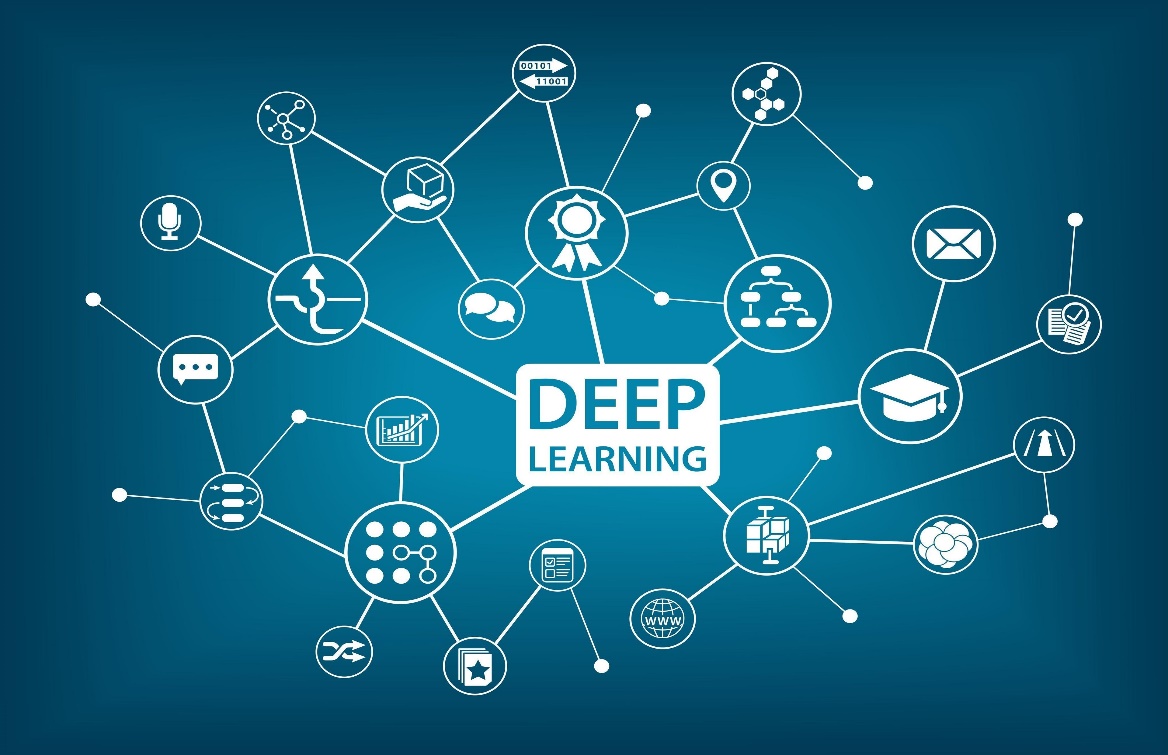
**Exploring advanced techniques like deep learning models (e.g., LSTM, BERT) for improved fake news detection accuracy.**

PHASE 2 SUBMISSION DOCUMENT

PROJECT:Fake News detection accuracy



**INTRODUCTION:**

In the modern era of computing, the news ecosystem has transformed from old traditional print media to social media outlets. Social media platforms allow us to consume news much faster, with less restricted editing results in the spread of fake news at an incredible pace and scale. In recent researches, many useful methods for fake news detection employ sequential neural networks to encode news content and social context-level information where the text sequence was analyzed in a unidirectional way. Therefore, a bidirectional training approach is a priority for modelling the relevant information of fake news that is capable of improving the classification performance with the ability to capture semantic and long-distance dependencies in sentences.

In this paper, we propose a BERT-based (Bidirectional Encoder Representations from Transformers) deep learning approach (FakeBERT) by combining different parallel blocks of the single-layer deep Convolutional Neural Network (CNN) having different kernel sizes and filters with the BERT. Such a combination is useful to handle ambiguity, which is the greatest challenge to natural language understanding. Classification results demonstrate that our proposed model (FakeBERT) outperforms the existing models with an accuracy of 98.90%.

**Content for project phase 2:**

consider exploring advanced techniques like deep learning models (e.g., LSTM, BERT) for improved fake news detection accuracy.

**Data source:**

In this phase, we can explore innovative techniques such as ensemble methods and deep learning architectures to improve the prediction system's accuracy and robustness.

**Dataset Link:**[**https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset**](https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset)

**About this file**

This dataset contains a list of articles considered as "fake" news

**Deep learning models:**

**What is Deep Learning?**

[Deep learning](https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-is-deep-learning) uses artificial neural networks to perform sophisticated computations on large amounts of data. It is a [type of machine learning](https://www.simplilearn.com/tutorials/machine-learning-tutorial/types-of-machine-learning) that works based on the structure and function of the human brain. Deep learning algorithms train machines by learning from examples. Industries such as health care, eCommerce, entertainment, and advertising commonly use deep learning.

### **Long Short Term Memory Networks (LSTMs)**

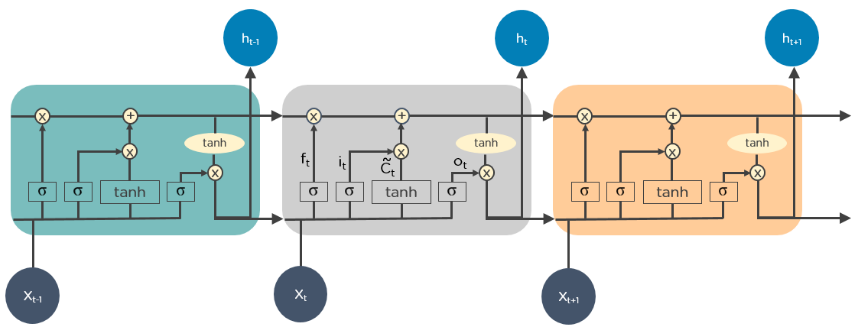
LSTMs are a type of Recurrent Neural Network (RNN) that can learn and memorize long-term dependencies. Recalling past information for long periods is the default behavior. LSTMs retain information over time. They are useful in time-series prediction because they remember previous inputs. LSTMs have a chain-like structure where four interacting layers communicate in a unique way. Besides time-series predictions, LSTMs are typically used for speech recognition, music composition, and pharmaceutical development.

**How Do LSTMs Work**

First, they forget irrelevant parts of the previous state

* Next, they selectively update the cell-state values
* Finally, the output of certain parts of the cell state

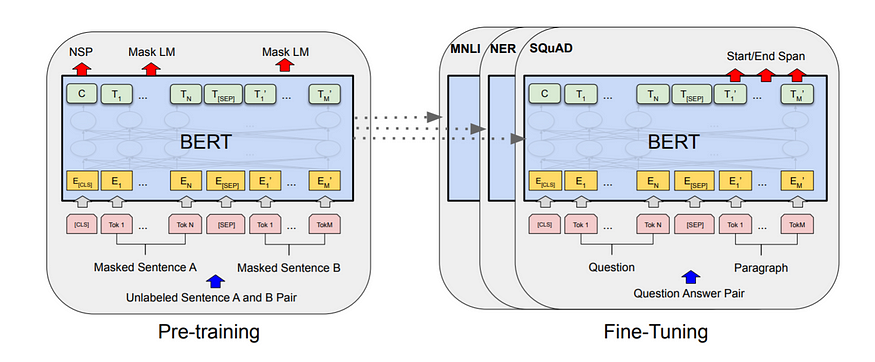
Below is a diagram of how LSTMs operate:



# **What is BERT?**

BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers and is a language representation model by Google. It uses two steps, pre-training and fine-tuning, to create state-of-the-art models for a wide range of tasks.

Its distinctive feature is the unified architecture across different downstream tasks — what these are, we will discuss soon. That means that the same pre-trained model can be fine-tuned for a variety of final tasks that might not be similar to the task model was trained on and give close to state-of-the-art results.



BERT consists of two steps. Source: [The paper](https://arxiv.org/abs/1810.04805).

As you can see, we first train the model on the pre-training tasks simultaneously. Once the pre-training is complete, the same model can be fine-tuned for a variety of downstream tasks. Note that a separate model is fine-tuned for a specific downstream task. So single pre-trained models can generate multiple downstream task specific models post fine tuning.

# **BERT Architecture**

Simply put, it is **a stack of Transformer’s Encoder**. You can read about Transformers in details in [my previous article](https://towardsdatascience.com/transformers-89034557de14). Or if you have some faint idea about it already, check out [this](https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/blocks/bert-encoder) absolutely bomb 3D diagram of the Encoder block used in BERT. **Seriously you can’t miss this!**

**Program:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import tensorflow as tf

from tensorflow.keras.preprocessing.text import Tokenizer

from tensorflow.keras.preprocessing.sequence import pad\_sequences

### **Load data**

In [3]:

fake\_df = pd.read\_csv('../input/fake-and-real-news-dataset/Fake.csv')

real\_df = pd.read\_csv('../input/fake-and-real-news-dataset/True.csv')

In [5]:

fake\_df.head()

Out[5]:

|  | title | Text | subject | date |
| --- | --- | --- | --- | --- |
| 0 | Donald Trump Sends Out Embarrassing New Year’... | Donald Trump just couldn t wish all Americans ... | News | December 31, 2017 |
| 1 | Drunk Bragging Trump Staffer Started Russian ... | House Intelligence Committee Chairman Devin Nu... | News | December 31, 2017 |
| 2 | Sheriff David Clarke Becomes An Internet Joke... | On Friday, it was revealed that former Milwauk... | News | December 30, 2017 |
| 3 | Trump Is So Obsessed He Even Has Obama’s Name... | On Christmas day, Donald Trump announced that ... | News | December 29, 2017 |
| 4 | Pope Francis Just Called Out Donald Trump Dur... | Pope Francis used his annual Christmas Day mes... | News | December 25, 2017 |

In [6]:

linkcode

real\_df.head()

| title | text | Subject | date |
| --- | --- | --- | --- |
| 0 | As U.S. budget fight looms, Republicans flip t... | WASHINGTON (Reuters) - The head of a conservat... | politicsNews | December 31, 2017 |
| 1 | U.S. military to accept transgender recruits o... | WASHINGTON (Reuters) - Transgender people will... | politicsNews | December 29, 2017 |
| 2 | Senior U.S. Republican senator: 'Let Mr. Muell... | WASHINGTON (Reuters) - The special counsel inv... | politicsNews | December 31, 2017 |
| 3 | FBI Russia probe helped by Australian diplomat... | WASHINGTON (Reuters) - Trump campaign adviser ... | politicsNews | December 30, 2017 |
| 4 | Trump wants Postal Service to charge 'much mor... | SEATTLE/WASHINGTON (Reuters) - President Donal... | politicsNews | December 29, 2017 |

**1.Count of Fake and Real Data**

In [9]:

print(data["target"].value\_counts())

fig, ax = plt.subplots(1,2, figsize=(19, 5))

g1 = sns.countplot(data.target,ax=ax[0],palette="pastel");

g1.set\_title("Count of real and fake data")

g1.set\_ylabel("Count")

g1.set\_xlabel("Target")

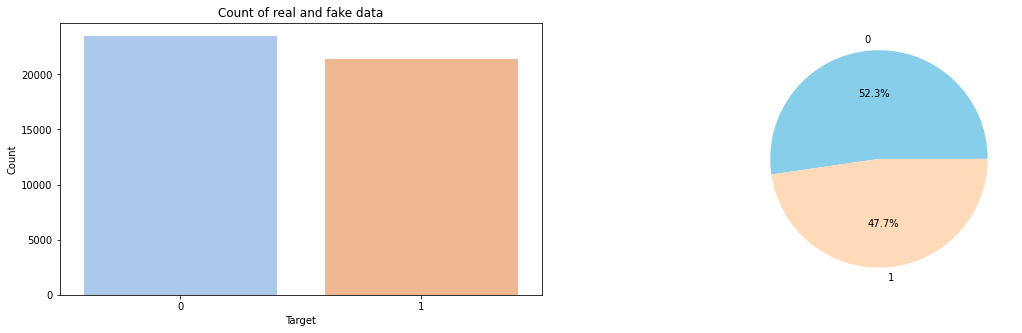
g2 = plt.pie(data["target"].value\_counts().values,explode=[0,0],labels=data.target.value\_counts().index, autopct='**%1.1f%%**',colors=['SkyBlue','PeachPuff'])

fig.show()

0 23481

1 21417

Name: target, dtype: int64



**2.Distribution of The Subject According to Real and Fake Data**

In [10]:

print(data.subject.value\_counts())

plt.figure(figsize=(10, 5))

ax = sns.countplot(x="subject", hue='target', data=data, palette="pastel")

plt.title("Distribution of The Subject According to Real and Fake Data")

politicsNews 11272

worldnews 10145

News 9050

politics 6841

left-news 4459

Government News 1570

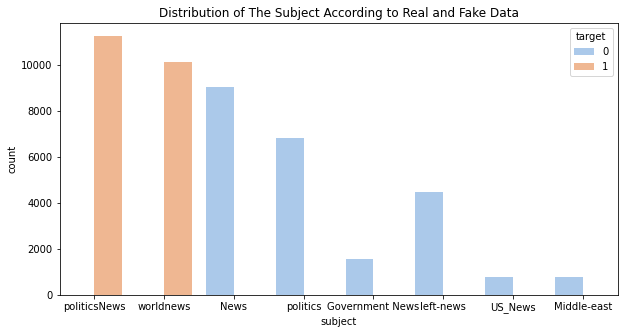
US\_News 783

Middle-east 778

Name: subject, dtype: int64

Out[10]:

Text(0.5, 1.0, 'Distribution of The Subject According to Real and Fake Data')



# **Data analysis, data visualization and creating new features**

In [12]:

*# Explore the target variable*

sns.countplot(x='fake', data=df\_news)

Out[12]:

<AxesSubplot:xlabel='fake', ylabel='count'>



In [13]:

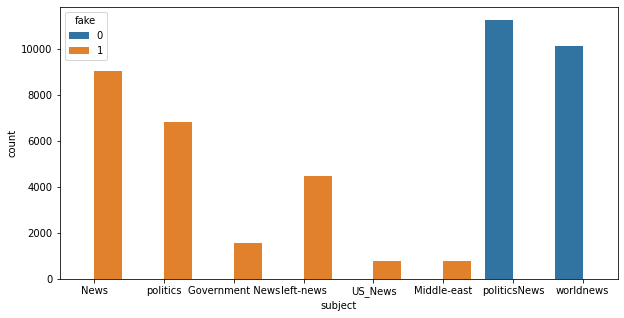
*# Explore the subject column*

plt.figure(figsize=(10,5))

sns.countplot(x='subject', data=df\_news, hue='fake')

Out[13]:

<AxesSubplot:xlabel='subject', ylabel='count'>



In [14]:

df\_news['date'] = pd.to\_datetime(df\_news['date'], errors='coerce') *# If 'coerce', then invalid parsing will be set as NaT.*

df\_news['Year'] = df\_news['date'].dt.year

df\_news['Month'] = df\_news['date'].dt.month

df\_news.head()

Out[14]:

|  | title | Text | subject | date | fake | Year | Month |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | Donald Trump Sends Out Embarrassing New Year’... | Donald Trump just couldn t wish all Americans ... | News | 2017-12-31 | 1 | 2017.0 | 12.0 |
| 1 | Drunk Bragging Trump Staffer Started Russian ... | House Intelligence Committee Chairman Devin Nu... | News | 2017-12-31 | 1 | 2017.0 | 12.0 |
| 2 | Sheriff David Clarke Becomes An Internet Joke... | On Friday, it was revealed that former Milwauk... | News | 2017-12-30 | 1 | 2017.0 | 12.0 |
| 3 | Trump Is So Obsessed He Even Has Obama’s Name... | On Christmas day, Donald Trump announced that ... | News | 2017-12-29 | 1 | 2017.0 | 12.0 |
| 4 | Pope Francis Just Called Out Donald Trump Dur... | Pope Francis used his annual Christmas Day mes... | News | 2017-12-25 | 1 | 2017.0 | 12.0 |

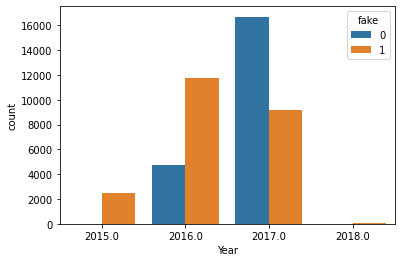
In [15]:

*# check the impact of year on the target*

sns.countplot(x='Year', data=df\_news, hue='fake')

Out[15]:

<AxesSubplot:xlabel='Year', ylabel='count'>



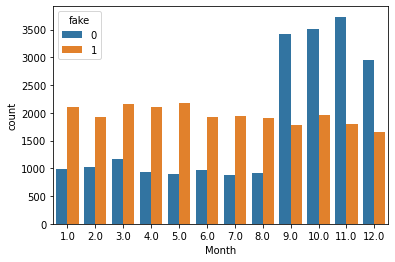
In [16]:

*# check the impact of month on the target*

sns.countplot(x='Month', data=df\_news, hue='fake')

Out[16]:

<AxesSubplot:xlabel='Month', ylabel='count'>



# **Conclusion:**

In this notebook, I used the TfidfVectorizer and PassiveAggressiveClassifier algorithms to detect "fake news" in the dataset. If you found these interesting, I highly encourage you to do further research yourself!

Thank you for reading until this far -- please upvote if you enjoyed my work and feel free to leave a comment and let me know your thoughts. All kinds of feedback & constructive criticism are appreciated!