**Deepfake Image Detection**

**Tool**

**Research Manual**

**by**

**Marcin Kaczmarek**

**C00242533**

**Supervisor: Richard Butler**

Carlow - 26/11/2021

# Abstract

Problem:

In today's world, any image, video or audio format can be changed. In addition, there may be multiple goals or intentions for doing so.

From the prosaic desire to beautify, improve the image to less innocent, such as trying to deceive someone or influence someone's opinion or perception of reality.

Objectives: Creating software that would allow assessing the examined image's authenticity and state whether it was manipulated or not.

Methods:

We are applying Benford's Law (Newcomb - Benford Law) to verify the authenticity of an image.

Table of Contents

[Abstract 2](#_Toc88859276)

[Introduction 4](#_Toc88859277)

[Deepfake 4](#_Toc88859278)

[Benefits of Deepfake Technology 5](#_Toc88859279)

[Methods of defence against Deepfake 5](#_Toc88859280)

[History 6](#_Toc88859281)

[Benford Law - Prime Digit Law 7](#_Toc88859282)

[Mathematical Explanation of Benford Law: 8](#_Toc88859283)

[Applications of the Benford distribution to data analysis 10](#_Toc88859284)

[Properties of sets subject to the Benford distribution 11](#_Toc88859285)

[Where can we apply Benford law? 12](#_Toc88859286)

[Fraud detection: 12](#_Toc88859287)

[Accuracy of mathematical models 13](#_Toc88859288)

[Analysis of the 2009 elections in Iran 13](#_Toc88859289)

[Combating deepfake 13](#_Toc88859290)

[Commercial Deepfake Detection Tool – Various Types 14](#_Toc88859291)

[FALdetector: 14](#_Toc88859292)

[Deepstar: 14](#_Toc88859293)

[Resemblyzer: 14](#_Toc88859294)

[Common image formats 14](#_Toc88859295)

[JPG format 14](#_Toc88859296)

[PNG 16](#_Toc88859297)

[GIF format 17](#_Toc88859298)

[TIFF Format 18](#_Toc88859299)

[RAW Format 19](#_Toc88859300)

[Image compression 20](#_Toc88859301)

[Project Technologies 25](#_Toc88859302)

[Alternative deepfake solutions 26](#_Toc88859303)

[NFT's 26](#_Toc88859304)

[Summary and Conclusions 28](#_Toc88859305)

[Table of Figures 29](#_Toc88859306)

[Table of Tables 29](#_Toc88859307)

[Bibliography: 30](#_Toc88859308)

# Introduction

This thesis is a project of the 4th year of studies in Cybercrime and I.T. Security.

This document will contain the research and steps to develop software to detect tampering with the tested images.

This research will be based on the Benford-Newcomb law, which is also a fascinating point of view on the world around us.

This work will refer to the phenomenon of deepfake that surrounds us in all kinds of media. This technology, with some benefits, has caused quite a stir.

It is no longer possible to detect the modifications made to images, video or audio at first glance. It's usually tricky even when using commercial deepfake detection tools, and they do not guarantee 100% originality.

My project will attempt to create a tool based on Benford prime number law and possibly perfectly suited to my project.

# Deepfake

Deepfake is a technology that allows generating realistically, confusingly similar to the original characters when used with digital voice modulation. - State-of-the-art artificial intelligence (A.I.) algorithms only need one photo and only a 5-second voice sample to be able to generate an almost identical deepfake hologram.

[1] (Botticello, 2020)

The first deepfake algorithms appeared at the end of 2017 and were used mainly in the porn industry. For example, a digitally generated face of a celebrity was added to the actor's body. As a result, we received a movie scene with a figure that resembled the Hollywood prototype. Today, the problem goes much deeper, and successfully fighting deepfake is more complex than ever.



Figure 1- Deepfake Example

[2] (ScreenRant, 2020)

A.I. is the future of technology, but it can pose a severe threat when in the wrong hands. The deepfake phenomenon is an excellent example of this because, with the use of artificial intelligence algorithms, it is possible to successfully manipulate selected groups, discredit individuals or even decide about the course of elections or the emergence of armed conflicts on a global scale.

However, deepfake is not only about politicians or celebrities. There are examples of it being a powerful weapon in the hands of cybercriminals, targeting profitable companies and influential entrepreneurs.

# Benefits of Deepfake Technology

Deepfake also has some positive aspects.

It is used in media, movies, games and many more.

In the movie industry, technology can be used to create movie scenes with dead actors. Make use of special effects and face edits. In addition, it allows to dub in any language.

# Methods of defence against Deepfake

Deepfake technologies can be tried to combat and detect using many programs such as:

* FALdetector
* Deepstar
* Visual DeepFake Detection
* DeepFake Audio Detection
* Resemblyzer

[3] (PenTestIT, 2020)

However, they do not give a hundred per cent certainty as a result.

They usually test the environmental change of the sampled image or small nuances of facial draws.

A little information can also give data contained in the multimedia files themselves, which you can examine if they were modified or not.

Element deciding on safety when using technology is the awareness of the threats that can be carried. Therefore, the more we know about the solutions used, the more scenarios of potential events we know, the more effective and reasonably we can counteract any dangers related to modern technologies.

[4] (Yu et al., 2021)

For main aspects that are subject to testing belong:

* Changes in face colours and improper lighting

Sometimes you can see with the naked eye some nuances in the colour of the skin and the facial features of the characters in the photos. Discolouration, inadequate lighting, and inadequate shading may indicate the likelihood of a false image.

* Reducing the number of eye blinks

When it comes to video footage, it is sometimes possible to detect unnatural eye movements, absent or unusual blinking, or surroundings of the face that do not match the rest.

* The unnatural appearance of human posture

Sometimes you can observe untypical human posture, distortion.

* Synchronisation of sound (if we speak about video files)

If the video conversion was carried out carelessly, you might notice a lack of synchronising the lip movements with the underlying sound.

* Blur or misalignment.

Image edges can sometimes be blurry or uneven. In addition, parts of the body can sometimes be in the wrong proportions, which can be easily seen in the photos.

[5] (us.norton.com, n.d.)

Fully effective tools for detecting deepfake algorithms have not yet been developed. Among the remedial measures that can be reached are technologies such as Sherlock AI, [6] which allows detecting anomalies in videos. Blockchain, which protects data registers from manipulation, or Natural Hash [7], enables content creators to embed in source materials. Non-editable digital watermarks. [8]

[6] (DeepQuanty, n.d.)

[7] (Du, Chen and Ke, 2018)

[8] (Locklizard, 2015)

## History

1. The Benford distribution is also called the Newcomb - Benford law, the law of odd numbers, or the first digit law.

While at the United States Naval Observatory Library, Newcomb noticed that the pages of the log tables were dirtier at the beginning and cleaner last pages. He concluded that those using logarithmic arrays are more likely to look for numbers that start with lower digits at the beginning of the collections. He published his discovery on the pages of the American Journal of Mathematics, but it was met with little enthusiasm, and his discovery was forgotten.

In 1938, Frank Benford, unaware of Newcomb's work, made the same discovery based on quality - the wear of pages in logarithmic tables. Fascinated by this phenomenon, Benford began to check whether his theory was found in other data collections, including river surfaces, numbers printed in newspapers, or even in prices. Finally, he presented the results of his research in an article published in Proceedings of the American Philosophical Society.

[9] (Gonsalves, 2020)

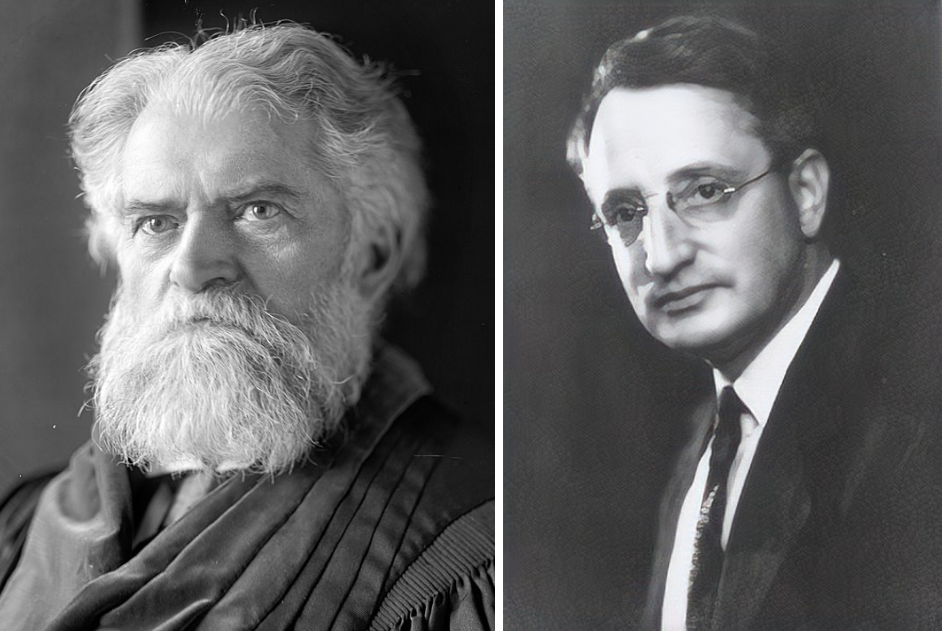


Figure 2- Simon Newcomb and Frank Benford

[9] (Gonsalves, 2020)

**Anecdote**

Let repeat a little anecdote here, unrelated to the subject of this study.

Has anyone heard of Stigler's law of eponymy?

It relates to the question of discoveries.

Discoveries are often named after someone else, someone who did not discover something first, and it is prevalent.

This law was proposed by the American statistics professor Stephen Stigler [10] in 1980 when he wrote that no scientific discovery was named after the original discoverer.

He also admitted that Stigler's Law itself was not his discovery but his predecessor, Robert Merton - who discovered it earlier.

[10] (Gizmodo, n.d.)

# Benford Law - Prime Digit Law

Who would have thought that the numbers surrounding us, such as bills, river paths from source to sea, the number of votes cast in elections, or the distance between cities, have common characteristics? So how can I say that seemingly random numbers can have commonalities?

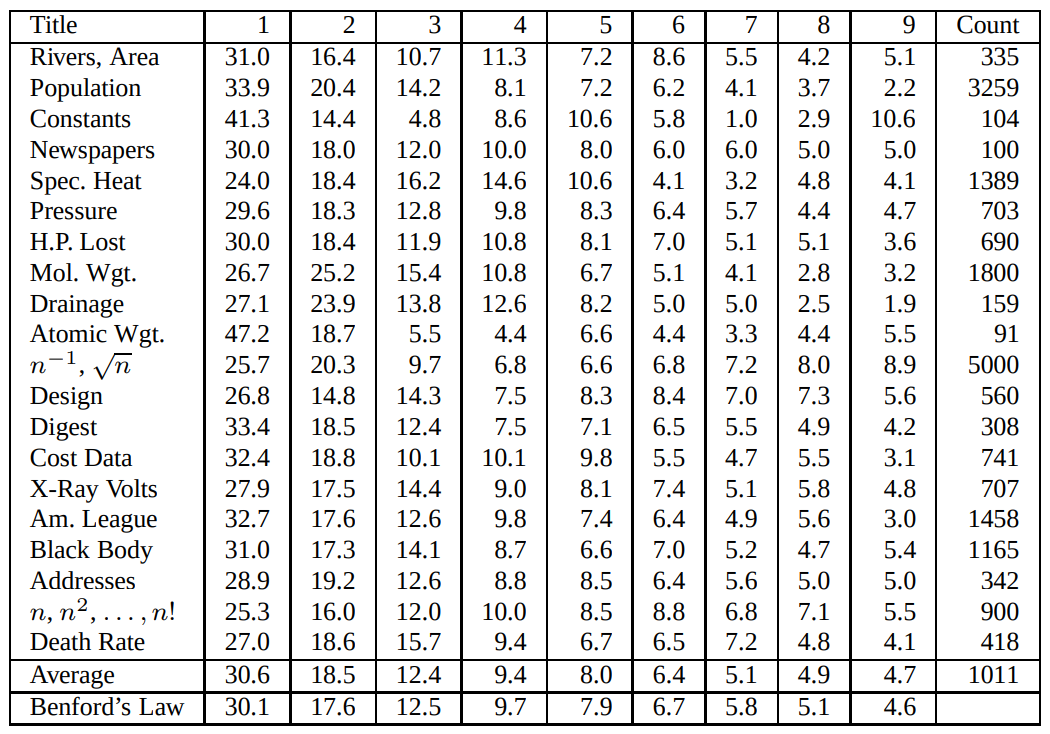


Table 1 Distribution of leading digits

[11] (Research & Development World, 2015)

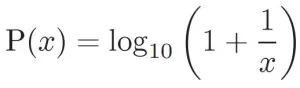
In the table above, we can see many sample datasets that Benford used to test his theory.

Seemingly unrelated sets of data - accurate data - make up numbers that reflect Benford's law.

The numbers in the last column at the bottom reflect the calculations used when applying Benford's law.

## Mathematical Explanation of Benford Law:

Thus, the world learned of an unusual regularity known as Benford's Law, Benford Distribution, or the Law of First (Significant) Numbers. The relationship describes the discrete Benford distribution.



[12] (Labs, 2016)

* x is the leading digit
* P (x) is the probability that x will be in the most significant position in the digit.

**Now something understandable:**

All datasets are made up of numbers, and numbers are the basis of math.

How often do we encounter the number "1" in a set of numbers?

Each number in the set from 1 to 9 is just a number.

So all these numbers should be the same percentage of our harvest.

That is:

1-11%

2 - 11%

3 - 11%

.

.

9-11%, right?

Not necessarily!

Benford found that these numbers appear in a completely different order - more likely to occur.

The number "1" occurs in about 30% of the cases, and the number "9" only in 5%.

Nevertheless, why is this happening?

Let us consider any set of numbers, for example, distance to school.

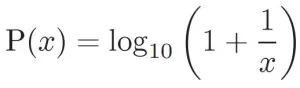
Some distances are short:1,2,3,4,5,6,7,8,9

and some longer:1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19 - look how much more we already have the number "1" in the first position!

Some distances are much longer: 21,22,23,24 ... 31,32,33,34 .... 41,42,45- where we have much more numbers such 2, 3, 4 and so on.

When the distances are long, such as 123,180,169, we have much more "1" in front.

So note that the number "1" appears much more frequently than 2, 3, 4, and so on up to 9.

**Benford law formula calculations:**

[12] (Labs, 2016)

**Calculation:**

|  |  |  |
| --- | --- | --- |
| P (1) = log10 (1 + 1/1)  = log10 (2)  = 0.30103  = **30.1** (rounded) | P (2) = log10 (1 + 1/2)  = log10 (1.5)  = 0.17609 ...  = **17.6**% (rounded) | P (3) = log10 (1 + 1/3)  = log10 (1.33333)  = 0.124938  = **12.5**% (rounded) |

And so on to the number 9.

**Distribution of probabilities:**

1 2 3 4 5 6 7 8 9

30.1% 17.6% 12.5% 9.7% 7.9% 6.7% 5.8% 5.1% 4.6%

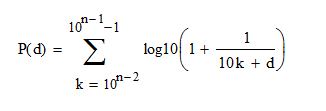
The approximate probability of individual digits in the most significant position is presented in the table below.



Table 2 Benford Law Distribution

The Benford distribution is used to validate figures, tax returns, or fraud. People typing numbers to appear random are not aware that specific numbers appear more often in the first position.

Benford also extended his law to 2nd, third and subsequent digits.



* d – is any digit from 1 to 9
* k – is a position of the digit

[13] (Invest Excel, 2012)

His calculations can be seen in the table below:

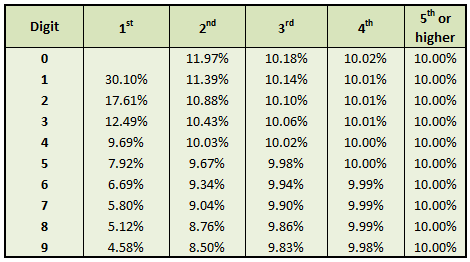


Table 3 The digit frequency predicted by equation

[13] (Invest Excel, 2012)

## Applications of the Benford distribution to data analysis

**Benford Distribution is used as a financial crime detection method**

* detecting false or incorrect accounting, recognising fraud tax,

Using the formula of Benford's law, you can trace the finances of a company (for example, expenses) and determine whether they are following Benford's law - that is, whether they are real and have not been modified [14]

[14] (Anon, n.d.)

* analysis of data from the stock exchange [14],

You can come across many instances of Benford's Law being used to control stock trading and investment returns. [15] One such example is the use of Benford's law as a tool to identify distortions and errors in financial statements used by the SEC (The State Examinations Commission).

[15] (Amiram, Bozanic and Rouen, 2015)

* analysis of prices auctioned at online auctions,

By using this law, you can confirm the authenticity of auctions carried out on popular Internet portals.

You can confirm the fairness of the auctions carried out or prove that the auctions are being manipulated.

[7] (Du, Chen and Ke, 2018)

**In information and technical sciences, this distribution is used for:**

* recognition of the authenticity of digitally saved photos,

In a word, detecting false and manipulated images is my project's goal.

[7] (Du, Chen and Ke, 2018)

**In the field of life sciences to:**

* assessing the effectiveness of drugs,

We can use Benford's law to confirm the authenticity and avoid errors when researching the effectiveness of the drugs used

[7] (Du, Chen and Ke, 2018)

* for testing the authenticity of messages on pollution levels

Do testing any other authentic datasets

And many other cases.

## Properties of sets subject to the Benford distribution

Not all large, natural (non-manipulated) data sets are subject to Benford's Law. Several necessary conditions must be met for the observation result to be credible.

When we undertake the use of Benford's Law to carry out the analysis, we need to consider several points.

Is the tested dataset compatible - suitable for testing?

Another thing is to consider what tests to run and how to interpret the results?

[16] (Durtschi, Hillison and Pacini, n.d.)

For example, most of the accounting data can be expected to follow the Benford distribution.

This is the result of the connection between numbers and, more strictly, with calculations that give specific results.

* Data must be collected in one dataset

Data collected for testing must be one dataset. [16]

* The dataset should not contain limits - we take into account the entire range of results without limitations.
* Datasets should not contain identifiers, telephone numbers or (artificially generated) bank accounts that do not conform to Benford's law. [16]
* Such datasets should contain lower-value numbers than the higher ones - for example, data from retail sales or river lengths. [16]

[16] (Durtschi, Hillison and Pacini, n.d.)

**Benford set attributes:**

* Invariability of the scale - multiplying all the values collected in our Benford set by a constant number, we get another new Benford set. [17]
* Base invariance - Benford's law applies to all number systems.
* Invariance of mathematical operations - all mathematical operations, such as addition and subtraction, exponentiation, multiplication and division of sets, form another Benford set, confirming the set's authenticity. [17]

[17] (Browne, 1998)

# Where can we apply Benford law?

### Fraud detection:

A leading example is the embezzlement of James Nelson, who was found guilty of the embezzlement of $ 1,878,687.58 in the town of Wayne (Arizona, USA) in 1992.

He did so by issuing 23 counterfeit checks. The fraud started with a small amount with successive amounts of fake reviews gradually increasing. Most of the checks were made for less than $ 100,000.

Detecting the crime was possible thanks to the breakdown of the first digits of the individual amounts.

The Benford Law Test is one of the first tests to be used in fraud detection! [20]

[18] (Invest Excel, 2012b)

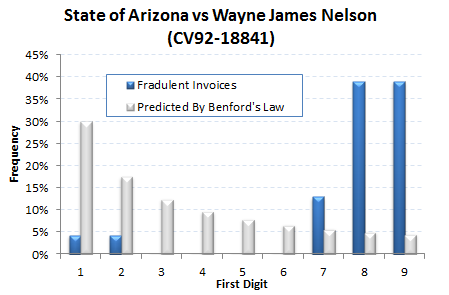


Table 4 Accounting fraud example [18] (Invest Excel, 2012b)

An excel sheet is a straightforward tool here.

It is enough to insert the tested data into the spreadsheet and evaluate the result with Benford's law.

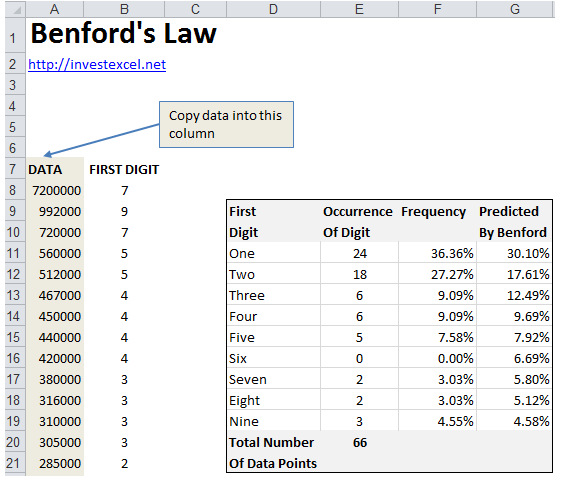


Table 5 An Excel sheet for Benford Law calculation

[18] (Invest Excel, 2012b)

### Accuracy of mathematical models

Benford's law is used to check the accuracy of mathematical models describing the evolution of data from various domains, for example, models of population change. Thus, we should obtain output that meets this relationship for input data that satisfies Benford's law. If this is not the case, it means that the model (algorithm) used disturbed the "natural" data distribution.

### Analysis of the 2009 elections in Iran

Boudewijn Roukema from the Nicolaus Copernicus University in Poland took a closer look at these elections.

He found inconsistencies in voting in some regions of the country with Benford's law.

In three of the six most significant regions, the current winner and president had far more votes than the rest.

It could have had a significant impact on the election results in Iran. [19]

[19] (Roukema, 2013)

# Combating deepfake

Probably the best solution is to use Anti-Deepfake technology. It may include:

1) tools to detect deepfake

2) content authentication

3) preventing the use of content to create deepfake. [20]

However, it is challenging to authenticate the content due to the number of files sent by the users.

It is a massive challenge for technology to go through published material in a short time.

As the development of anti-deepfake technology is not enough. Organisations must also adopt these technologies; for example, any government in any country may have ideas to regulate the use of deepfake. [20]

[20] (Westerlund, 2019)

# Commercial Deepfake Detection Tool – Various Types

### FALdetector:

The open-source tool was written in Python to help detect faces changed in Photoshop.

It detects image distortions applied to human faces, implementing a fully trained model with false images generated automatically using Photoshop software scripts. [21]

This software focuses on one particular type of manipulation in Photoshop – image distortion applied to a face. It is a commonly known and used modification of human faces.

The main task of this software is to classify and predict whether the face has changed, as well as to identify places where this may have occurred, and to apply a mechanism that allows these changes to be reversed. [21]

[21] (Anon, n.d.)

### Deepstar:

Deepstar is a Python-based open-source toolbox based on artificial intelligence that helps detect deepfake videos.

It includes many different anti-counterfeit detection algorithms, such as automatic detection of facial manipulation in movies or focusing on the lips of an object.

Thanks to this, it is possible to judge with a high degree of probability whether the examined character was manipulated or not. [22]

[22] (PenTestIT, 2020a)

### Resemblyzer:

The Resemblyzer uses a voice encoder and a deep learning model.

It creates the characteristics of the tested voice, creating its detailed description taking into account 256 values characterising the human voice.

It is perfect for verifying the speaker, verifying whether the voice has been modified or not, and many other uses. [23]

[23] (Jemine, n.d.)

# Common image formats

To implement the project, it is necessary to decide on the file format that will be the most appropriate and valuable.

There are several formats for saving photos, the most popular of which are JPEG, GIF, PNG, and TIFF. Each of them has its advantages.

## JPG format

JPG, or Joint Photographic Experts Group, or JPEG, is perhaps the best-known image format. It is the default option for most photos saved as it is photo-friendly, thanks to the virtually unlimited colour of the display. [24]

JPG also offers the ability to choose a degree of image compression from 0 per cent (high compression) to 100 per cent (no compression). Most designers choose anything in the range of 60 to 70 per cent. Images still look good at this compression level, but file sizes are much smaller. [24]

JPG uses lossy compression and does not preserve the original data when compressed. Therefore, every time a photo is saved again and exported as a JPG file, it will degrade.

The JPG format is most often used for photos, photography, and anything with many colours. [25]

**JPEG benefits**

* Great for high colour and photography
* Easily reduce file sizes
* Renders consistently in email clients

**JPEG negatives**

* No transparency
* Creates jagged edges on the text
* No animation
* No automatic metadata for search must contain information.

[24] (Iqbal, 2019)

[25] (Rishab2612, 2016)



Figure 3 JPEG example image

[26] (FixThePhoto.com, n.d.)

In JPEG format, can compress files - shrink an image to take up less memory. The downside to compression is the loss of image quality. Each time compress a photo in JPEG format, the conversion algorithm removes some pixels, and the process is irreversible. [27]

[27] (DISCRETE COSINE TRANSFORM DISCRETE COSINE TRANSFORM FOR VARIOUS VALUES OF U AND V, n.d.)

**This is crucial information for our deepfake files research.**

## PNG

PNG offers something JPG cannot do - transparency.

PNG, or Portable Network Graphics, is a format designed for the Internet.

PNG is so popular because it transfers elements like logos to website designs.

There are two types of PNG files - PNG-8 and PNG-24. PNG-8 uses a more limited colour palette with only 256 colours, has slightly better transparency, and exports small. PNG-24 uses an unlimited colour palette, retains transparency but exports at a larger size. Both types of PNG have lossless compression. [28]

Although PNG formats are similar to GIF files, they do not support animation. Therefore, this format is most commonly used for icons, small photos, or any image that requires transparency.

[28] (W3.org, 2019)

**PNG benefits**

* It supports transparency
* Suitable for images with text
* PNG formats render logos well
* Includes embedded text description for search engines
* PNG-8 has a small file size and is the most lightweight
* Exports without jagged edges

**PNG negatives**

* File sizes overgrow for large files such as images
* Some older email clients have trouble rendering them
* No animation
* PNG-24 files can be considerable; not so suitable for web sharing.

[28] (W3.org, 2019)



Figure 4 PNG example image

[26] (FixThePhoto.com, n.d.)

## GIF format

GIF or Graphical Interchange Format is a highly compressed image.

A GIF image is typically up to 8 bits per pixel and 256 colours throughout the image. Unlike JPEG, an image that can display up to 16 million colours and touches the boundaries of the human eye as far as possible.

Animated GIF combines multiple images or frames into a single file and displays them in sequence to generate an animated clip or short movie. [29]

GIF file format

Conceptually, GIF files have a fixed-size graphic area filled with zero or more images. Some GIF files divide the graphics area or fixed-size blocks into sub-pictures that can act as animated frames for an animated GIF. The GIF format uses a pixel depth of 1 to 8 bits to store bitmap data. The RGB colour model and palette data are always used to store the images. [29]

**Animated GIFs benefits**

* Small File Size
* Professional Looking Images
* Convey Messages Better

**Animated GIFs negatives**

* Limited Colour Pattern
* Editing Is Not Possible
* Internet Connection Matters

[29] (Ayres, 2017)

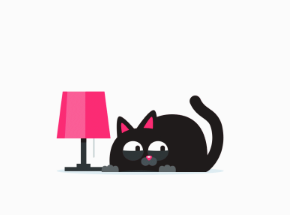


Figure 5 GIF example image

[26] (FixThePhoto.com, n.d.)

To see, please click this link:

<https://i.pinimg.com/originals/aa/18/c2/aa18c28f148884017b56e3e3e61f6e71.gif>

## TIFF Format

TIFF, or less commonly TIF, stands for "Tagged Image File Format".

TIFF is a standard format for photography and print data transfer. All objects, including vector and text information, are stored as raster data in a TIFF file. [30]

"Raster Data" - A method of defining 2D graphical data with black and white or coloured dots instead of vector data. The resolution of raster data is usually determined by the number of pixels or dots/inches.

The TIFF file supports grayscale as well as RBG, CMYK, and LAB colour spaces. The format allows a colour depth of up to 16 bits per colour channel and is ideal for data exchange during RAW conversion. [30]

**TIFF benefits**

* Lossless compression
* High image quality
* Ideal for photograph printouts
* Transparencies and layers

**TIFF negatives**

* Not compatible with many browsers
* Requires more storage space

[30] (Iqbal, 2019b)



Figure 6 TIFF example image

[26] (FixThePhoto.com, n.d.)

## RAW Format

RAW is a format that contains unique data in the electronics of cameras or scanners such as colour, white balance and exposure. This data is beneficial for photo editing, which allows you to obtain images of the highest quality. [31]

The RAW format requires many memory cards and the ability to use software such as Photoshop and Lightroom before they can be used. However, this gives you equal editing capability and perfect image quality. [31]

**RAW benefits**

* image of the highest quality
* great image editing capabilities

**RAW negatives**

* necessary resources such as memory, software and processing time

[31] (dpbestflow.org, n.d.)



Figure 7 RAW example image

[26] (FixThePhoto.com, n.d.)

**Summary of the discussed image formats**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | JPG | TIFF | PNG | GIF | RAW |
| Lossless Compression | YES | YES | YES | YES | YES |
| Lossy Compression | YES | YES | YES | YES | YES |
| Transparency Support | NO | YES | YES | NO | NO |
| File size per 4K frame, RGB | 4 MB | 27 M.B. | 10.5 MB | 24MB | 9 MB |
| Support for 4K | YES | YES | YES | YES | YES |
| Maximum Supported bit depth | 12 | 32 | 8 | 8 | 32 |
| Support for floating-point math | NO | YES | YES | NO | NO |

Table 6 Image formats summary

**The most helpful format for my project**

**JPEG**

Considering a suitable image format, which would be the most practical and valuable for the project, I decided to focus on the JPEG format. [32]

JPEG is probably the most widely used photo format we deal with today. This format is also an excellent source of information. The technical JPEG markers can therefore be used for counterfeit detection at the fundamental level.

These tags (metadata) are detailed information about the algorithms used and the devices or software modifying the image and information on the device and conditions during image capture. The JPEG format can be compressed many times without losing quality. On the one hand, this loss can be a negative feature, but from another point of view, one can tell if the file has been compressed and approximately how many times.

JPEG is based on the DCT (Discrete cosine transform) technique. [32]

DCT is the process of splitting an image into different frequency parts. First, the minor fundamental frequencies are rejected. Moreover, the critical frequencies are retained so that the image can be used in the decompression process.

However, the reconstructed image may contain some distortions.

[32] (unix4lyfe.org, n.d.)

# Image compression

**What is Compression?**

Compression is a process that reduces the size of a digital image by reducing the amount of information contained in the image.

These data are beneficial if we want to edit the image and make professional processing. Still, if we do not need them, they are unnecessary and can be deleted without losing the image quality. [33]

This compression allows images to be transferred more efficiently and also reduces disk space for storage.

Technically, it is the process of converting a two-dimensional array into a data set, which for my project will be the most optimal form of testing the originality of the examined images. [33]

Unlike compression, decompression is the process of decoding an image into its source version.

[33] (Rishab2612, 2016b)

**JPEG compression process**

JPEG compression is the key to image verification, and it uses DCT (cosine functions) to retrieve the data.

For this purpose, several different algorithms can be used, but they all follow the same methodology.

At the outset, there is a **transformation**: the elimination of unnecessary information contained in our image.

The next step is **quantisation**, which allows the loss of some data and leaves the necessary ones, which provides our image with the best possible quality. [33]

**Entropy coding** is the next step in the application of various algorithms to reduce the image size. Entropy coding is the transformation of the input information to the corresponding variable-length prefix codeword. [33]

[33] (Rishab2612, 2016b)

1. The first step is to transform the colours to **the luminance-chrominance model**.

If the saved image is in RGB format, then each part is in the range of numbers.

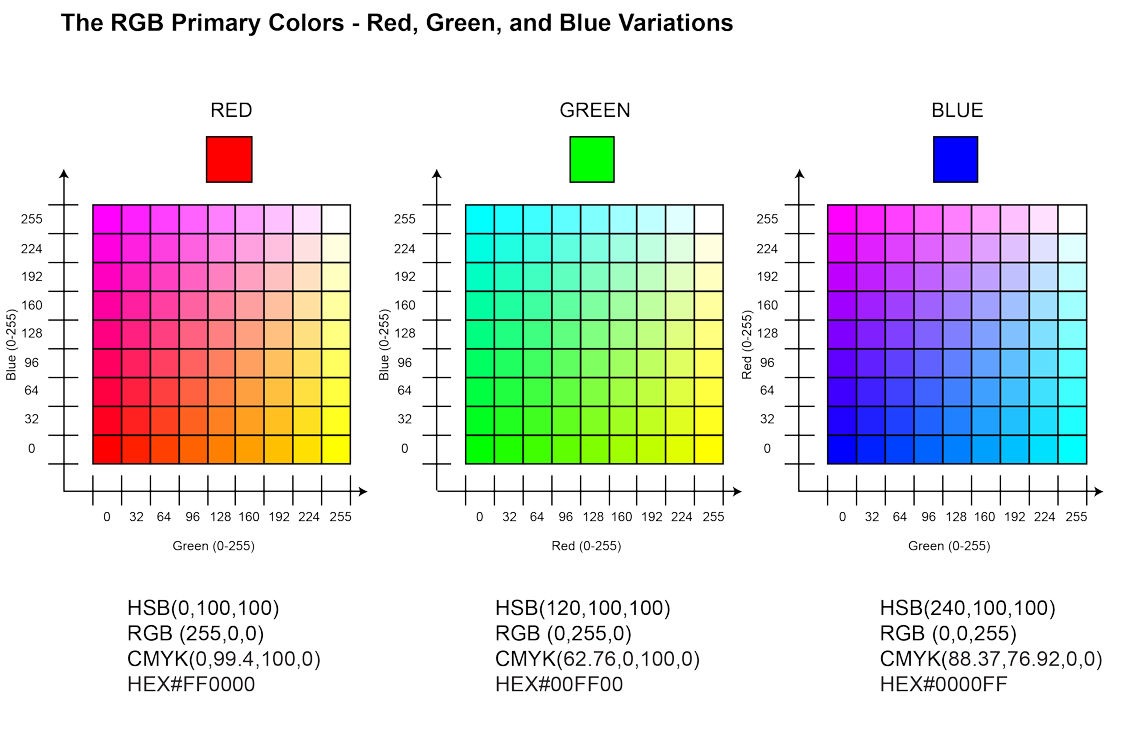


Figure 8 RGB colour ranges

1. In the second step, the image is split into 8 × 8 blocks of pixel size.

Each component creates a matrix of the sizes indicated above, and then, to facilitate the calculations, the value 128 is added to each element of the matrix. This will allow us to create a matrix where all parts will be in the range (-128 to 127). [33]

[33] (Rishab2612, 2016c)

1. **A discrete two-dimensional cosine transform (DCT)** is computed for each block.

As a result of this calculation, we will obtain a matrix with actual values corresponding to the average values and the frequency of changes within the block. [34]

[34] (Scribd, n.d.)

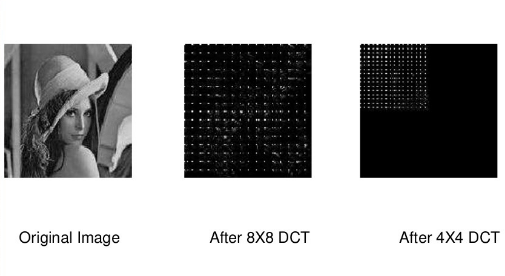


Figure 9 Compression steps

The discrete cosine transform (DCT) is designed to decorate the data contained in the image. And after its completion, each correlation coefficient can be coded separately.

As mentioned before, DCT removes unnecessary information. [34]

[34] (Scribd, n.d.)

1. At this stage, we already have **the cosine transform of the data block**, which then undergoes quantisation - i.e. the removal of non-existent data from the point of view of image quality.

The constants obtained in the quantisation process now form the quantisation matrix.

After the rounded values, many of them will be zero, which translates into the compressed image's size. [34]

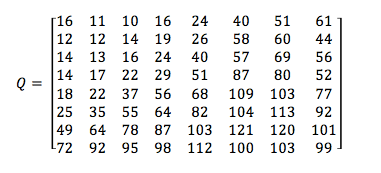


Figure 10 8x8 matrix [34]

[34] (Scribd, n.d.)

1. The next step is to prepare the data for encoding.

For this, our matrix is converted to a zig-zag algorithm presented below.

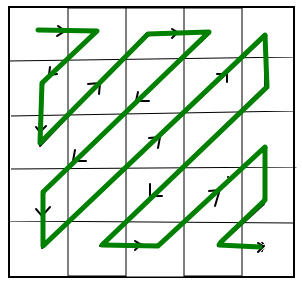


Figure 11 Zig-Zag method [35]

The last step is **entropy coding**, replacing the most common elements with words – codes.

In the end, we receive a stream of binary data.

Quantisation often gives a coefficient equal to zero, which can be successfully removed in the process and thus, compression takes place. [35]

[35] (M, 2016)

DCT and quantisation process example:

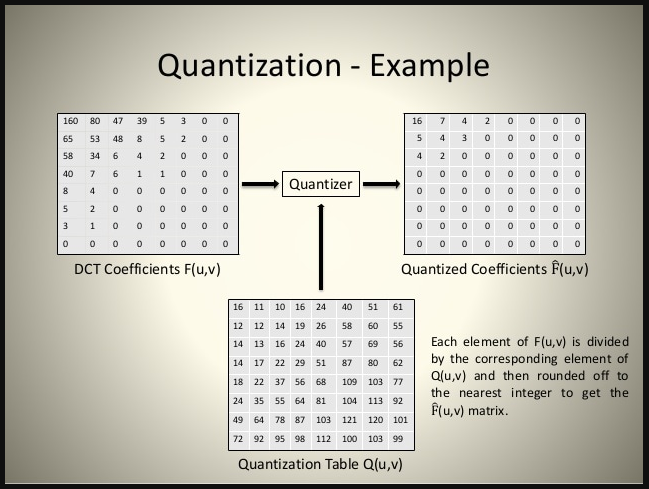


Figure 12 DCT Quantization [ 35]

[35] (M, 2016)

Here we have an original, a luminance-chrominance model

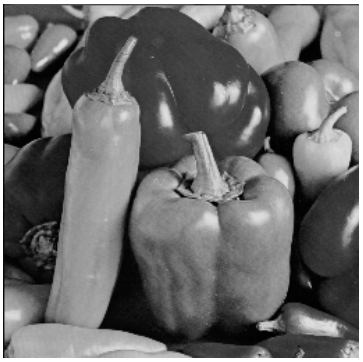
****

Figure 13 Original picture example [36]

and here, after DCT processing:

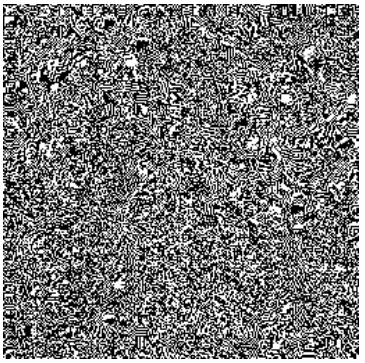


Figure 14 Picture during DCT processing [36]

Each block 8x8 is quantised using a quantisation matrix, and its size is reduced.

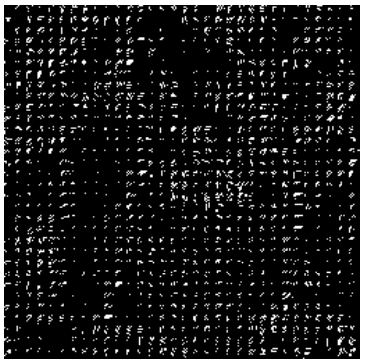
****

Figure 15 Quantised DCT of the picture. [36]

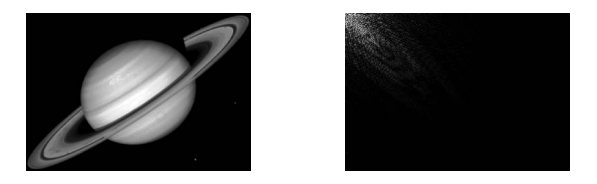
****

Figure 16 Original and processed image [37]

[37] (Syed and Khayam, 2003)

# Project Technologies

At this point, it should be mentioned what the correlation between Benford's law and JPEG file compression is.

Well, file compression, due to its nature and the way it is performed, can produce an excellent set of data that could be used in testing Benford's law.

The received data can be implemented in **Python** and tested for possible violations of this law or not.

Python seems quite well equipped with tools and databases to help implement this project as a development environment.

It has many libraries with mathematical databases, including Benford's Law. [38]

[38] (Mata, n.d.)

The functionality of my tool can be embedded in a web application that provides easy access from any type of operating system and any translator.

To design and build this web application, you may be tempted to use **the Flask framework.** [39]

Flask is a Python web platform built with a small core and easy to extend philosophy initially designed and developed by Armin Ronacher in 2010 and has become the most popular Python web platform in 2018. [40]

The Flask web framework is a code library that speeds up and simplifies web application development by providing common patterns for building reliable, scalable, and easy-to-maintain web applications. Since the turn of the century, it has been common practice in web development projects to use an existing web framework, except in very unusual situations. In addition to Flask, popular web frameworks include Django, ASP.NET, Laravel, and Ruby on Rails.

[39] (flask.palletsprojects.com, n.d.)

[40] (says, 2021)

Web frameworks provide code functionality or extensions to perform everyday operations required to run web applications.

These joint operations include:

* URL Routing
* Support and validation of input forms
* HTML, XML, JSON and other output formats with the template engine Database connection configuration and persistent data manipulation with object-relational mapping (ORM)
* Internet security against inter-site request forgery (CSRF), SQL injection, cross-site scripting (XSS) and other common malicious attacks
* Session storage and recovery [39]

[39] (flask.palletsprojects.com, n.d.)

Flask is a web framework, not a web programming language. Therefore, Flask expects Python files which Flask will then process.

# Alternative deepfake solutions

We have been struggling with deepfake for a long time, and it does not seem easy. It would be unambiguous. At first glance, someone could say whether a given image is original or not. You could pay attention to the technology that occupies an increasingly significant field in the arena of file authentication, whatever they may be - be it photos, works of art or otherwise.

## NFT's

Why am I even mentioning NFT's here, and what exactly is that?!?

NFT, or Non-Fungible Token, is a non-exchangeable, technically unchanged, unhackable certificate. It confirms the uniqueness of digital items in the blockchain register, i.e., a system used to store and transmit information about transactions concluded on the Internet.

NFT lets you see who owns a virtual item: file, picture, GIF, song, video, meme, etc.

It confirms their originality like a painter's signature on a canvas. It can also be one copy of many, for example, collectable virtual cards.

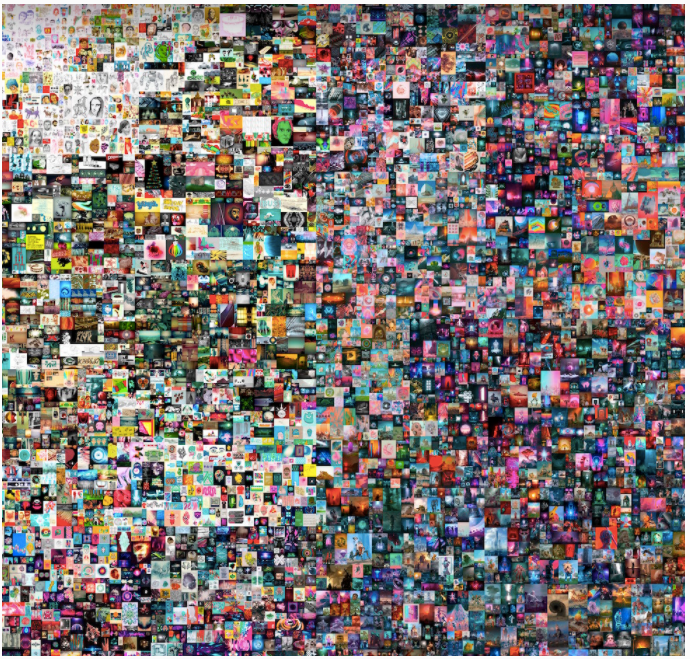


Figure 17 First and most popular NFT's [41]

One of the first and most expensive NFT works to date in history, sold for $ 69 million by Beeple - Mike Winkleman.

[41] (www.christies.com, n.d.)



Figure 18 A modern NFT image example

[42] (Torres, 2021)

**But how does this relate to my project?**

I mention this because blockchain and NFT's technology can be perfect for the authentication of all files, including image files.

Thanks to this, it would be possible to eliminate deepfake, and it would take even a few seconds for an inexperienced user to determine whether something is original or not. [42]

You would end up having trouble deciding what is genuine and what is not. Whether something has been modified or not.

In a word, the era of deepfake could be over once and for all. [43]

[43] (BeInCrypto, 2021)

[44] (NFT Plazas, n.d.)

# Summary and Conclusions

This project aims to create a tool that can verify if a tested JPEG image has been modified or not. For example, has anyone manipulated it, and is it possible to say with certainty its actual state when it comes to authenticity?

As a finished project, I will create a web application built on the Flask framework and Python as the most appropriate language for my application.

# Table of Figures

[Figure 1- Deepfake Example 5](#_Toc88859162)

[Figure 2- Simon Newcomb and Frank Benford 7](#_Toc88859163)

[Figure 3 JPEG example image 16](#_Toc88859164)

[Figure 4 PNG example image 17](#_Toc88859165)

[Figure 5 GIF example image 18](#_Toc88859166)

[Figure 6 TIFF example image 19](#_Toc88859167)

[Figure 7 RAW example image 20](#_Toc88859168)

[Figure 8 RGB colour ranges 22](#_Toc88859169)

[Figure 9 Compression steps 23](#_Toc88859170)

[Figure 10 8x8 matrix [34] 23](#_Toc88859171)

[Figure 11 Zig-Zag method [35] 24](#_Toc88859172)

[Figure 12 DCT Quantization [ 35] 24](#_Toc88859173)

[Figure 13 Original picture example [36] 25](#_Toc88859174)

[Figure 14 Picture during DCT processing [36] 25](#_Toc88859175)

[Figure 15 Quantised DCT of the picture. [36] 25](#_Toc88859176)

[Figure 16 Original and processed image [37] 26](#_Toc88859177)

[Figure 17 First and most popular NFT's [41] 27](#_Toc88859178)

[Figure 18 A modern NFT image example 28](#_Toc88859179)

# Table of Tables

[Table 1 Distribution of leading digits 8](#_Toc88858793)

[Table 2 Benford Law Distribution 10](#_Toc88858794)

[Table 3 The digit frequency predicted by equation 11](#_Toc88858795)

[Table 4 Accounting fraud example [18] (Invest Excel, 2012b) 13](#_Toc88858796)

[Table 5 An Excel sheet for Benford Law calculation 14](#_Toc88858797)

[Table 6 Image formats summary 21](#_Toc88858798)

# Bibliography:

[1]

Botticello, C. (2020). *What is a Deepfake?* [online] Deepfake. Available at: <https://medium.com/deepfake/what-is-a-deepfake-9bcc72f3eb1e> [Accessed 25 Nov. 2021].

[2]

ScreenRant. (2020). *How Deepfake Technology Works*. [online] Available at: https://screenrant.com/deepfake-videos-explained-how/ [Accessed 25 Nov. 2021].

[3]

PenTestIT. (2020b). *List of Open Source Deepfake Detection Tools*. [online] Available at: https://pentestit.com/open-source-deepfake-detection-tool-list/ [Accessed 25 Nov. 2021].

[4]

Yu, P., Xia, Z., Fei, J. and Lu, Y. (2021). A Survey on Deepfake Video Detection. *IET Biometrics*. / [Accessed 26 Nov. 2021].

[5]

us.norton.com. (n.d.). *How to spot deepfake videos — 15 signs to watch for*. [online] Available at: https://us.norton.com/internetsecurity-emerging-threats-how-to-spot-deepfakes.html. / [Accessed 26 Nov. 2021].

[6]

DeepQuanty. (n.d.). *Sherlock AI*. [online] Available at: http://deepquantyailabs.com/sherlock-ai/ [Accessed 25 Nov. 2021].

[7]

Du, L., Chen, Z. and Ke, Y. (2018). Image Hashing for Tamper Detection with Multiview Embedding and Perceptual Saliency. *Advances in Multimedia*, [online] 2018, p.e4235268. Available at: https://www.hindawi.com/journals/am/2018/4235268/ [Accessed 25 Nov. 2021].

[8]

Locklizard. (2015). *Document Watermarks: Digital Watermarking Security for Documents, Files - Locklizard*. [online] Available at: https://www.locklizard.com/document-watermarking/ / [Accessed 26 Nov. 2021].

[9]

Gonsalves, R.A. (2020). Benford's Law — A Simple Explanation. *Medium*. [online] 6 Oct. Available at: https://towardsdatascience.com/benfords-law-a-simple-explanation-341e17abbe75. / [Accessed 26 Nov. 2021].

[10]

Gizmodo. (n.d.). *Stigler's Law: Why nothing in science is ever named after its actual discoverer*. [online] Available at: https://gizmodo.com/stiglers-law-why-nothing-in-science-is-ever-named-afte-5820736 [Accessed 25 Nov. 2021].

[11]

Research & Development World. (2015). *How a Simple Observation from the 1800s about Patterns in Big Data Sets Can Fight Fraud*. [online] Available at: https://www.rdworldonline.com/how-a-simple-observation-from-the-1800s-about-patterns-in-big-data-sets-can-fight-fraud/ [Accessed 25 Nov. 2021].

[12]

Labs, Fsp. (2016). *Forensics and Benford's Law*. [online] Event Log Explorer blog. Available at: https://eventlogxp.com/blog/forensics-and-benfords-law/ [Accessed 25 Nov. 2021].

[13]

Invest Excel. (2012a). *Find Financial Fraud with Benford's Law*. [online] Available at: https://investexcel.net/benfords-law-excel [Accessed 25 Nov. 2021].

[14]

Anon, (n.d.). [online] Available at: https://www.acfe.com/uploadedFiles/Shared\_Content/Products/Self-Study\_CPE/UsingBenfordsLaw\_2018\_final\_extract.pdf [Accessed 26 Nov. 2021].

[15]

Amiram, D., Bozanic, Z. and Rouen, E. (2015). Financial statement errors: evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, 20(4), pp.1540–1593. / [Accessed 26 Nov. 2021].

[16]

Durtschi, D., Hillison, W. and Pacini, C. (n.d.). *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data*. [online] Available at: https://www.researchgate.net/publication/241401706\_The\_Effective\_Use\_of\_Benford’s\_Law\_to\_Assist\_in\_Detecting\_Fraud\_in\_Accounting\_Data [Accessed 26 Nov. 2021].

[17]

Browne, M. (1998). *Following Benford's Law, or Looking Out for No. 1*. [online] Available at: https://people.math.harvard.edu/~ctm/home/text/class/harvard/154/11/html/media/benfords\_law.pdf [Accessed 26 Nov. 2021].

[18]

Invest Excel. (2012). *Find Financial Fraud with Benford's Law*. [online] Available at: https://investexcel.net/benfords-law-excel/ [Accessed 26 Nov. 2021].

[19]

Roukema, B.F. (2013). A first-digit anomaly in the 2009 Iranian presidential election. *Journal of Applied Statistics*, 41(1), pp.164–199. / [Accessed 26 Nov. 2021].

[20]

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, [online] 9(11). Available at: https://timreview.ca/article/1282. / [Accessed 26 Nov. 2021].

[21]

Anon, (n.d.). *Detecting Photoshopped Faces by Scripting Photoshop*. [online] Available at: https://arxiv.org/pdf/1906.05856.pdf [Accessed 26 Nov. 2021].

[22]

PenTestIT. (2020). *Deepstar: An Open Source Deepfake Detection Toolkit*. [online] Available at: https://pentestit.com/deepstar-open-source-deepfake-detection-toolkit/ [Accessed 26 Nov. 2021].

[23]

Jemine, C. (n.d.). *Resemblyzer 0.1.1.dev0: Analyse and compare voices with deep learning*. [online] PyPI. Available at: https://pypi.org/project/Resemblyzer/ [Accessed 26 Nov. 2021].

[24]

Iqbal, K. (2019a). *JPEG - Image File Format*. [online] docs.fileformat.com. Available at: https://docs.fileformat.com/image/jpeg [Accessed 26 Nov. 2021].

[25]

Rishab2612 (2016a). *Comparison between JPEG(DCT) and JPEG 2000(DWT) compression standards*. [online] Available at: https://www.slideshare.net/Rishab2612/comparison-between-jpegdct-and-jpeg-2000dwt-compression-standards [Accessed 26 Nov. 2021].

[26]

FixThePhoto.com. (n.d.). *What Is Highest Quality Image Format?* [online] Available at: https://fixthephoto.com/highest-quality-image-format.html. / [Accessed 26 Nov. 2021].

[27]

DISCRETE COSINE TRANSFORM DISCRETE COSINE TRANSFORM FOR VARIOUS VALUES OF U AND V. (n.d.). [online] Available at: https://d1b10bmlvqabco.cloudfront.net/attach/iizifk1n4713c/idqzrrsbms75iv/inczelhjcc5k/image\_seminars.pdf [Accessed 26 Nov. 2021].

[28]

W3.org. (2019). *Portable Network Graphics (PNG) Specification (Second Edition)*. [online] Available at: https://www.w3.org/TR/2003/REC-PNG-20031110/./ [Accessed 26 Nov. 2021].

[29]

Ayres, C. (2017). *6 Advantages and Disadvantages of Animated Gifs*. [online] ConnectUS. Available at: https://connectusfund.org/6-advantages-and-disadvantages-of-animated-gifs. / [Accessed 26 Nov. 2021].

[30]

Iqbal, K. (2019). *TIFF - Image File Format*. [online] docs.fileformat.com. Available at: https://docs.fileformat.com/image/tiff [Accessed 26 Nov. 2021].

[31]

dpbestflow.org. (n.d.). *Raw File Formats | dpBestflow*. [online] Available at: https://dpbestflow.org/file-format/raw-file-formats [Accessed 26 Nov. 2021].

[32]

unix4lyfe.org. (n.d.). *Discrete Cosine Transform*. [online] Available at: https://unix4lyfe.org/dct-1d [Accessed 26 Nov. 2021].

[33]

Rishab2612 (2016a). *Comparison between JPEG(DCT) and JPEG 2000(DWT) compression standards*. [online] Available at: https://www.slideshare.net/Rishab2612/comparison-between-jpegdct-and-jpeg-2000dwt-compression-standards. / [Accessed 26 Nov. 2021].

[34]

Scribd. (n.d.). *DCT | PDF | Data Compression | Matrix (Mathematics)*. [online] Available at: https://www.scribd.com/document/79004855/dct [Accessed 26 Nov. 2021].

[35]

M, A.K. (2016). *JPEG Image Compression*. [online] Slideshare.net. Available at: https://www.slideshare.net/AishwaryaKM1/jpeg-image-compression-56894348. / [Accessed 26 Nov. 2021].

[36]

Anon, (n.d.). *,PDJH &RPSUHVVLRQ DQG WKH 'LVFUHWH &RVLQH 7UDQVIRUP*. [online] Available at: https://www.math.cuhk.edu.hk/~lmlui/dct.pdf [Accessed 26 Nov. 2021].

[37]

Syed, A. and Khayam (2003). *The Discrete Cosine Transform (DCT): Theory and Application 1*. [online] Available at: https://www.egr.msu.edu/waves/people/Ali\_files/DCT\_TR802.pdf. / [Accessed 26 Nov. 2021].

[38]

Mata, R. (n.d.). *benford-law: A python tool to calculate the Benford's Law from a dataset or image*. [online] PyPI. Available at: https://pypi.org/project/benford-law/ [Accessed 26 Nov. 2021].

[39]

flask.palletsprojects.com. (n.d.). *Welcome to Flask — Flask Documentation (2.0.x)*. [online] Available at: https://flask.palletsprojects.com/en/2.0.x/./ [Accessed 26 Nov. 2021].

[40]

says, D.F. (2021). *Top 10 Best Web Development Technologies to use in 2021 | ISHIR- Software Development Company India*. [online] ISHIR | Software Development India. Available at: https://www.ishir.com/blog/9821/top-10-best-web-development-technologies-to-use-in-2021.htm. / [Accessed 26 Nov. 2021].

[41]

www.christies.com. (n.d.). *Beeple's masterwork: the first purely digital artwork offered at Christie's | Christie's*. [online] Available at: https://www.christies.com/features/Monumental-collage-by-Beeple-is-first-purely-digital-artwork-NFT-to-come-to-auction-11510-7.aspx. / [Accessed 26 Nov. 2021].

[42]

Torres, L. (2021). *The Trevor Lawrence x adidas NFT Collection Will Support the Jacksonville Community | SoleSavy News*. [online] SoleSavy. Available at: https://solesavy.com/trevor-lawrence-adidas-nft-collection-information/ [Accessed 26 Nov. 2021].

[43]

BeInCrypto. (2021). *NFTs Explained: What are NFTs and How do they Work?* [online] Available at: https://beincrypto.com/learn/nfts-explainer [Accessed 26 Nov. 2021].

[44]

NFT Plazas. (n.d.). *What Are NFTs? - A Simple Explanation For Beginners*. [online] Available at: https://nftplazas.com/what-are-nfts [Accessed 26 Nov. 2021].