BTC Price Movement Prediction

Group 17

StockUpz

Table of Content

Problem Statement

How may we predict the price movement of cryptocurrency based on various information?

In other words...

Whether the price of BTC will rise or fall after a given day

Datasets

Yahoo Finance

via yfinance

BTC-USD
S&P 500
US Interest Rates

Kaggle

Web scraped

Crypto News Articles

Mendeley Data

Yazd University

Crypto Tweets from 50 Experts

Approach

1 Exploration

Attempt to study impact of different information on BTC prices

2 Shortlist

Shortlist the most impactful models for consideration

3 Selection

Final selection for final model

1 Exploration

Attempt to study impact of different information on BTC prices

Experimentation:

Output Calculation

$$\text{Output for Day 1} = \frac{\text{BTC-USD Closing Price Day 2}}{\text{BTC-USD Closing Price Day 1}}$$

1 Exploration

Experimentation:

Try different combinations of features

S&P-Interest Rates

Technical Indicators

News

Tweets

1

Exploration

Technical Indicators

What are technical indicators?

Using BTC-USD Data:

RSI_14 day

Speed and magnitude of price change

MACD

Moving average convergence divergence

12-26 days periods

Signal strength of trend or trend reversals)

20-Day Bollinger Bands

Volatility

Close - Open

High - Low

1

Exploration

Cryptocurrency Technical Indicators

Automated Feature Selection and Grid Search

1. Select n best features using SelectKBest

2. Pass these n features into the model, conduct grid search with cross-validation, select the best hyperparameters based on aggregated f1 score over k-folds

3. For the best estimator , print its hyperparameters ,features selected as well as scores

4. Repeat for different n (3 to max_number of feature columns in dataset)

Models Used

1. Adaboost
2. Random Forest
3. SVM (Poly & RBF Kernel)

1

Exploration

Cryptocurrency Technical Indicators

Top 3 models

Model	**Selected features**	**Hyperparameter**	**Backtest F1 score**
SVM (Poly kernel)	'Open', 'High', 'Close', 'Adj Close', 'Volume', 'RSI_14', 'macd', 'macd_signal', 'BB_middle', 'BB_lower', 'Close-Open', 'High-Low', 'OBV', 'BB_height	{'svm__C': 10, 'svm__coef0': 1.0, 'svm__degree': 4}	**0.70**
SVM (RBF kernel)	'Open', 'High', 'Close', 'Adj Close', 'Volume', 'RSI_14', 'macd', 'macd_signal', 'BB_lower', 'High-Low', 'OBV', 'BB_height'	{'svm__C': 0.1, 'svm__coef0': 0.0, 'svm__degree': 4}	0.60
Random Forest	'Open', 'High', 'Close', 'Adj Close', 'Volume', 'RSI_14', 'macd', 'macd_signal', 'BB_upper', 'BB_middle', 'BB_lower', 'Close-Open', 'High-Low', 'OBV', 'BB_height'	{'RF__max_depth': 10, 'RF__min_samples_leaf': 2, 'RF__min_samples_split': 5, 'RF__n_estimators': 50}	0.58

1

Exploration

News and BTC-USD

Approach

Try to predict the price movement based on news sentiment and polarity

Extend prediction to using News article

Models Used

1. Logistic Regression
2. Neural Network

Dataset merged

	Date	Close	price_change	next_day_close_increased	sentiment_polarity	sentiment_subjectivity
0	2021-10-12	56041.058594	-1360.039062	0	0.16	0.50
1	2021-10-15	61593.949219	701.769531	1	0.00	0.00
2	2021-10-18	62026.078125	-2235.914063	0	0.14	0.45
3	2021-10-19	64261.992188	-1730.843750	0	0.10	0.40
4	2021-10-27	58482.386719	-2139.750000	0	0.00	0.00
...	...	...	...	...	...	...
30963	2023-12-18	42623.539063	353.011719	1	0.00	0.00
30964	2023-12-18	42623.539063	353.011719	1	0.00	0.10
30965	2023-12-18	42623.539063	353.011719	1	-0.40	0.69
30966	2023-12-18	42623.539063	353.011719	1	-0.10	0.37
30967	2023-12-18	42623.539063	353.011719	1	0.20	0.53

30968 rows × 6 columns

1

Exploration

News and BTC-USD

Models Used

Logistic Regression with Sentiment

Selected features:

Sentiment_polarity| Sentiment_subjectivity

Output:

Class of 0 or 1

- Performed undersampling and oversampling

Test Accuracy: 0.48

Test Classification Report:

	precision	recall	f1-score	support
0	0.61	0.44	0.51	959
1	0.37	0.55	0.44	590
accuracy			0.48	1549
macro avg	0.49	0.49	0.48	1549
weighted avg	0.52	0.48	0.48	1549

BackTesting Results for Logistic Regression Model

Test Accuracy: 0.37

Test Classification Report:

	precision	recall	f1-score	support
-1	0.00	0.00	0.00	1
0	0.44	0.47	0.46	17
1	0.25	0.25	0.25	12
accuracy			0.37	30
macro avg	0.23	0.24	0.24	30
weighted avg	0.35	0.37	0.36	30

1

Exploration

News and BTC-USD

Models Used

Simple Neural Network

Selected features:

Sentiment_polarity| Sentiment_subjectivity

Output:

class of 0 or 1

Layers :

- Linear
- Dropout
- Relu
- sigmoid

	precision	recall	f1-score	support
0	0.00	0.00	0.00	902
1	0.42	1.00	0.59	646
accuracy			0.42	1548
macro avg	0.21	0.50	0.29	1548
weighted avg	0.17	0.42	0.25	1548

```
Epoch [1/15], Validation Loss: 0.7030
Epoch Accuracy: 0.4173
Epoch [2/15], Validation Loss: 0.7070
Epoch Accuracy: 0.4173
Epoch [3/15], Validation Loss: 0.6961
Epoch Accuracy: 0.4173
Epoch [4/15], Validation Loss: 0.7044
Epoch Accuracy: 0.4173
Epoch [5/15], Validation Loss: 0.7064
Epoch Accuracy: 0.4173
Epoch [6/15], Validation Loss: 0.6927
Epoch Accuracy: 0.5827
Epoch [7/15], Validation Loss: 0.7124
Epoch Accuracy: 0.4173
Epoch [8/15], Validation Loss: 0.7096
Epoch Accuracy: 0.4173
Epoch [9/15], Validation Loss: 0.7090
Epoch Accuracy: 0.4173
Epoch [10/15], Validation Loss: 0.7135
Epoch Accuracy: 0.4173
Epoch [11/15], Validation Loss: 0.7089
Epoch Accuracy: 0.4173
Epoch [12/15], Validation Loss: 0.7089
Epoch Accuracy: 0.4173
Epoch [13/15], Validation Loss: 0.7061
Epoch Accuracy: 0.4173
Epoch [14/15], Validation Loss: 0.7013
Epoch Accuracy: 0.4173
Epoch [15/15], Validation Loss: 0.7147
Epoch Accuracy: 0.4173
```

```
Test Accuracy: 0.3809
```

1

Exploration

Tweets

Approach

Try to predict the price movement based on Tweets about cryptocurrency.

Also consider how influential the Tweeter is and the volume of BTC traded that day

1. Selected features:
Tweets | Significant_coeff | BTC_Vol

2. Tokenise Tweets using Keras tokenizer

3. Combine the tokenised tweets with normalised signficant_coeff and BTC_vol

1

Exploration

Tweets

Models Used

Sequence Classification with LSTM Recurrent Neural

Model Result

Train-Test

	precision	recall	f1-score	support
0	0.98	1.00	0.99	3044
1	0.00	0.00	0.00	63
accuracy			0.98	3107
macro avg	0.49	0.50	0.49	3107
weighted avg	0.96	0.98	0.97	3107

Back testing

	precision	recall	f1-score	support
-1	0.00	0.00	0.00	0
0	1.00	0.57	0.72	30
1	0.00	0.00	0.00	0
accuracy			0.57	30
macro avg	0.33	0.19	0.24	30
weighted avg	1.00	0.57	0.72	30

1

Exploration

S&P-Interest Rates

Approach

Predict next day price movement based on the S&P 500 index data & interest rate data, combining with the BTC-USD dataset.

Models Used

1. LSTM
2. SVM (Poly & RBF Kernel)

Steps

1. Data Pre-processing
   - Dates where market is closed, forward filled data
   - MinMax Scale
2. Data Sequences
   - 10-days windows (short term prediction and price is volatile, trial and error)
   - Prevent data leakage
3. Train-Test
4. Find Threshold for Classification

1

Exploration

S&P-Interest Rates

Model Used

LSTM

Model Result

Train-Test

Classification Report:

	precision	recall	f1-score	support
0	0.55	0.20	0.29	30
1	0.60	0.88	0.71	41
accuracy			0.59	71
macro avg	0.57	0.54	0.50	71
weighted avg	0.58	0.59	0.54	71

Back testing

Classification Report:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	14
1	0.36	1.00	0.53	8
accuracy			0.36	22
macro avg	0.18	0.50	0.27	22
weighted avg	0.13	0.36	0.19	22

1

Exploration

S&P-Interest Rates

Approach

Predict next day price movement based on the S&P 500 index data & interest rate data, combining with the BTC-USD dataset.

Models Used

1. LSTM
2. SVM (Poly & RBF Kernel)

Steps

1. Data Pre-processing

2. Train-test Split
    - 80% of fold size for training
    - 20% of fold size for testing

3. Perfrom grid search with cross validation
    - K-folds = 10

1

Exploration

S&P-Interest Rates

Backtesting F1 Score with the best parameters

SVM (Poly Kernel)

	precision	recall	f1-score	support
0	0.57	1.00	0.73	12
1	0.00	0.00	0.00	9
accuracy			0.57	21
macro avg	0.29	0.50	0.36	21
weighted avg	0.33	0.57	0.42	21

SVM (RBF Kernel)

	precision	recall	f1-score	support
0	0.53	0.67	0.59	12
1	0.33	0.22	0.27	9
accuracy			0.48	21
macro avg	0.43	0.44	0.43	21
weighted avg	0.45	0.48	0.45	21

2 Shortlist

Shortlist the most impactful models for consideration

2

Shortlist

Overview of all Models

BTC-USD

0.58 Random Forest

0.70 SVM (Poly Kernel)

0.60 SVM (RBF)

NEWS

0.38 Logistic Regression

0.38 Neural Network

0.57 LSTM

TWEETS

0.57 LSTM

S&P-INTEREST RATES

0.36 LSTM

0.57 SVM

3 Selection

Final selection for final model

3

Selection

Stacking

Output of one model is the input for the next

Combining

Aggregate relevant features to be inputted into a single model

3

Selection

RATIONALE

Test if combining outputs from individual models would give better results

Stacking

Output of one model is the input for the next

Model	Number of features	Selected features	Hyperparameter	K-fold	Backtesting F1 score
SVM (Poly kernel)	3	'predictions_BTC_only', 'prediction_Tweet_only', 'prediction_News_only’	{'svm__C': 10, 'svm__coef0': 2.0, 'svm__degree': 3]	5	**0.42**
SVM (RBF kernel)	3	'predictions_BTC_only', 'prediction_Tweet_only', 'prediction_News_only’	{'svm__C': 10, 'svm__gamma': 100}	5	**0.42**
RF	3	'predictions_BTC_only', 'prediction_Tweet_only', 'prediction_News_only’	{'RF__max_depth': None, 'RF__min_samples_leaf': 1, 'RF__min_samples_split': 2, 'RF__n_estimators': 50}	5	**0.42**
Adaboost	3	'predictions_BTC_only', 'prediction_Tweet_only', 'prediction_News_only’	{'adaboost__learning_rate': 0.1, 'adaboost__n_estimators': 50}	5	**0.42**

3

Selection

Combining

Aggregate relevant features to be inputted into a single model

RATIONALE

Test if using raw inputs from all datasets would be better since stacking did not exactly work out

Model	Number of features	Selected features	Hyperparameter	K-fold	Backtesting F1 score
SVM (Poly kernel)	24	['Open', 'High', 'Close', 'Adj Close', 'Volume', 'RSI_14', 'macd', 'macd_signal', 'BB_upper', 'BB_middle', 'BB_lower', 'Close-Open', 'High-Low', 'OBV', 'BB_height', 'sentiment_polarity', 'sentiment_subjectivity', 'Interest Rate', 'SP_Open', 'SP_High', 'SP_Low', 'SP_Close', 'SP_AdjClose', 'SP_Volume']	{'svm__C': 10, 'svm__coef0': 2.0, 'svm__degree': 3}	5	**0.57**
AdaBoost	8	['Open', 'Close', 'Adj Close', 'macd', 'BB_middle', 'High-Low', 'OBV', 'sentiment_subjectivity']	{'adaboost__learning_rate': 0.5, 'adaboost__n_estimators': 150}	10	**0.57**
Random Forest	3	['Open', 'macd', 'BB_middle']	{'RF__max_depth': 10, 'RF__min_samples_leaf': 4, 'RF__min_samples_split': 5, 'RF__n_estimators': 100}	5	**0.53**

Final Model

We decide to evaluate which is the best based on F1 score on back tested set for a fair comparison. This model will be used in our GUI

Best model:

Model	Selected features	Hyperparameter	Backtest F1 score
SVM (Poly kernel)	'Open', 'High', 'Close', 'Adj Close', 'Volume', 'RSI_14', 'macd', 'macd_signal', 'BB_middle', 'BB_lower', 'Close-Open', 'High-Low', 'OBV', 'BB_height	{'svm__C': 10, 'svm__coef0': 1.0, 'svm__degree': 4}	**0.70**

GUI Demo

Thank you

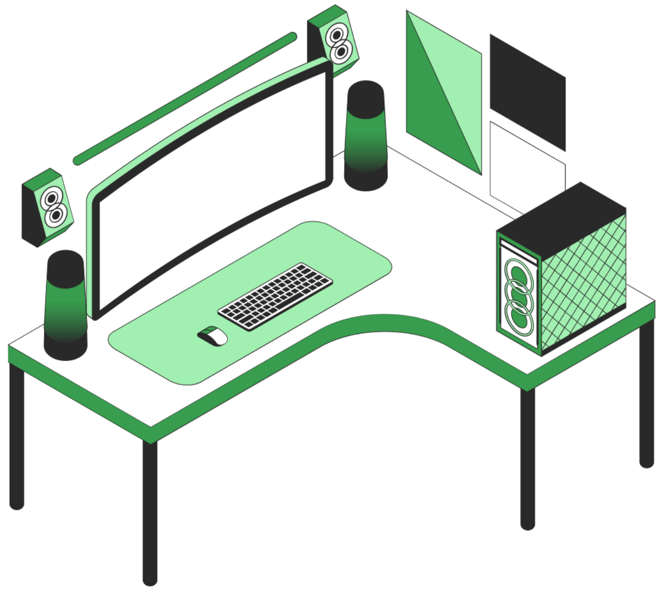
Group 17

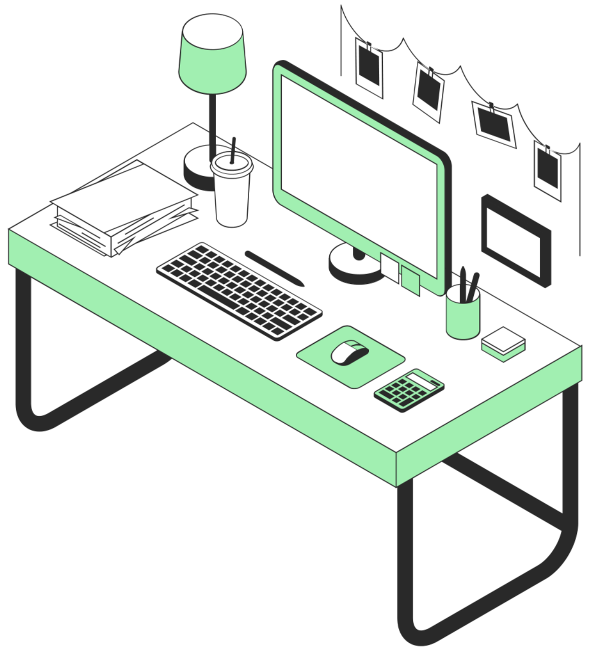
StockUpz

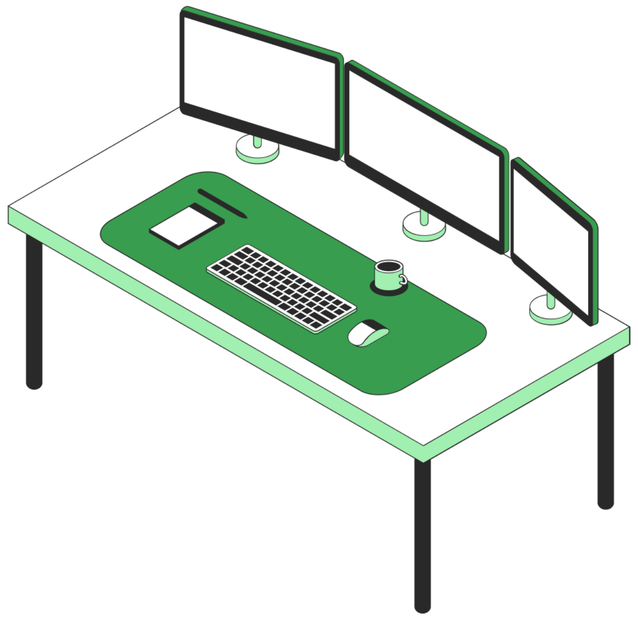
Resource Page

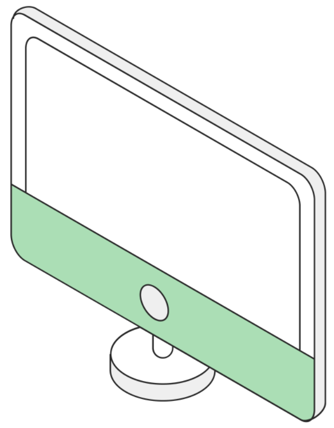
Use these design resources in your Canva Presentation. Happy designing!

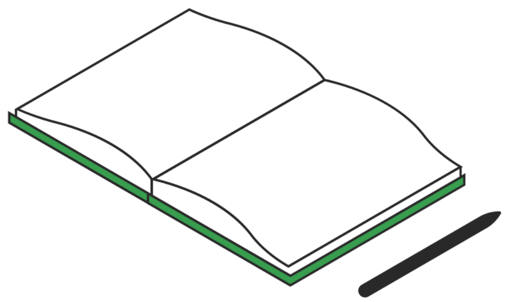
Don't forget to delete or hide this page before presenting.

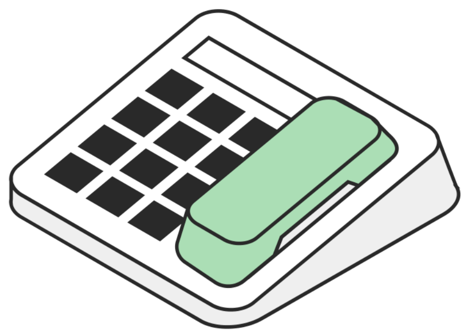








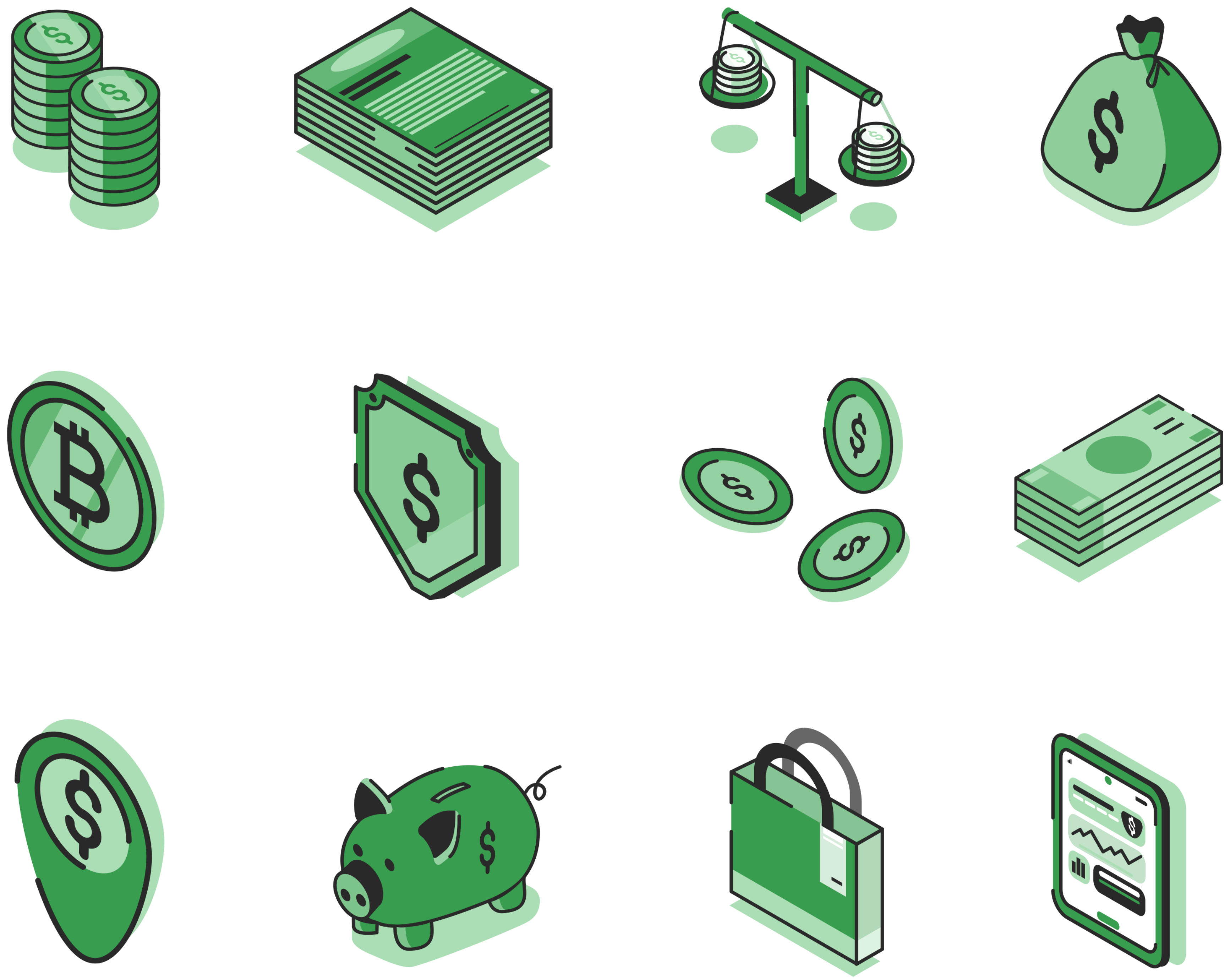




Resource Page

Use these design resources in your Canva Presentation. Happy designing!

Don't forget to delete or hide this page before presenting.



Resource Page

Find the magic and fun in presenting with Canva Presentations. Press the following keys while on Present mode!

Delete or hide this page before presenting.

B for blur	C for confetti
D for a drumroll	M for mic drop
O for bubbles	Q for quiet
U for unveil	Any number from 0-9 for a timer