W U T I S

Algorithmic Trading Division

Peer Group Investing Automated by ChatGPT

Karina Pekarek-Kostka, Daniel Eder, Kirill Gusev, Valentin Ennser, Anna Siniaeva

Vienna, June 28th, 2023

Team Overview

Algorithmic Trading

Karina Pekarek-Kostka

Head of Algorithmic Trading

- Task distribution
- Data cleaning
- ChatGPT algorithm

- MSc. (WU) – Incoming
- BSc. (WU)

Daniel Eder

Associate Project Head

- Project idea
- Optimization
- Strategy

- BSc. (WU) – 6th Sem.

Kirill Gusev

Associate Project Supervisor

- Data cleaning
- Scientific perspective

- MSc. (UW) – 4th Sem.
- BSc. (UW)

Valentin Ennser

Analyst

- Clustering model
- Errors fixing

- BSc. (WU) – 4th Sem.
- BSc. (TU) – 6th Sem.

Anna Siniaeva

Analyst

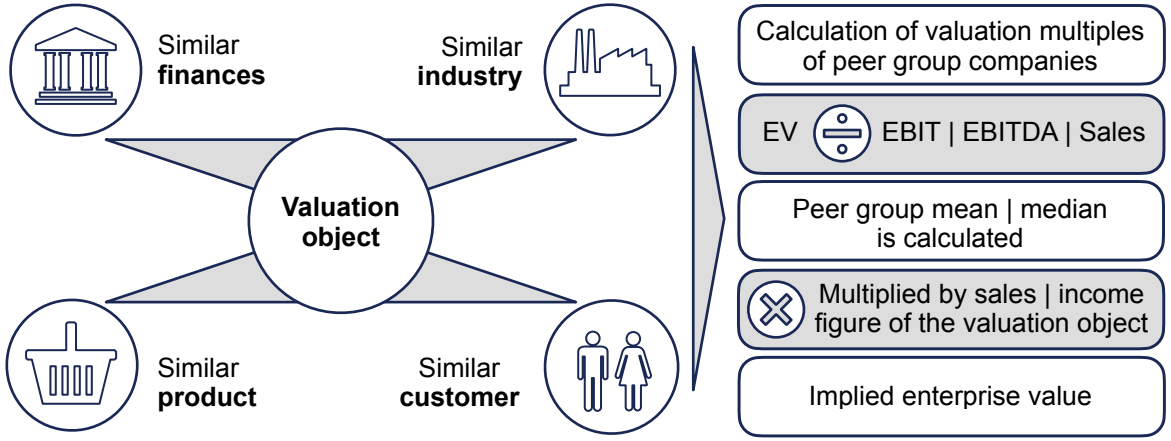
- Clustering model
- Storyline

- MA. (LBS) – 4th Sem.
- BSc. (MSU)

Problem statement

Two novel approaches should tackle the implicit limitations of manual peer group valuation

Overview of Comparative Company Analysis



Automatic peer groups forming with Clustering and ChatGPT approaches

Industry classification (165 industries)

Predefined set of internal financial metrics

Predefined set of market variables

Potential week points of manual valuation

- Time-consuming
- Highly unwieldy
- Might be biased

1 **CLUSTERING**

2

PEER GROUP | MANUALLY

PEER GROUP | AUTOMATICALLY

Geertsema, P. G. & Lu, H. (2021) Relative Valuation with Machine Learning. *Journal of Accounting Research, Forthcoming*. http://dx.doi.org/10.2139/ssrn.3740270

Data used for modeling

The data from 4,000 companies were used to build algorithms

W U T I S

Data collected from various sources

TOTAL DATA
~4,000 companies

TEST SET
~1,000 companies

TRAIN SET
~3,000 companies

20 QUARTERS | 5 YEARS

~80,000 observations

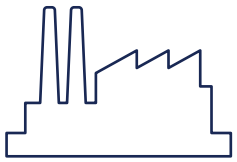
Faced biases

1. Non relevant variables were removed from the data set
2. Companies with N/A's in vital financial data were removed from the dataset
3. The preset time horizon of 5 years

Available and input variables

DATA FOR EACH COMPANY

COMPANY CATEGORY



- Entity Name
- Entity ID
- Primary Industry
- Global Region
- Quarter of the Year
- 1st Level Primary Industry
- 2d Level Primary Industry
- Exchange
- Ticker

FINANCIAL VARIABLES

- EBIT
- EBITDA
- Net Debt
- Total Debt
- Net Income
- Total Liabilities
- Cash
- Total Revenue
- Total Assets

RATIOS

- Market Cap
- EBIT Margin
- Profit Margin
- Leverage
- EV/EBIT
- EV
- EV/EBITDA
- EV/Sales
- EBITDA Margin
- Debt/Assets
- Cash/Assets
- P/E
- beta

Clustering model

K-means clustering was used for forming the peer groups

K-means clustering: definition and model

K-MEANS CLUSTERING

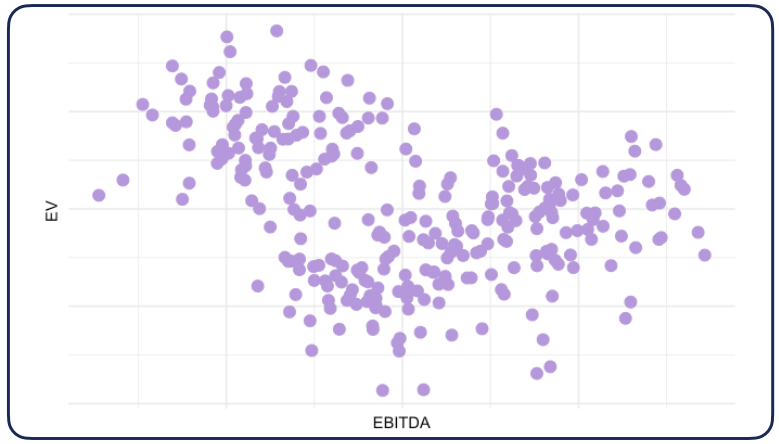
Focuses on the distances between observations

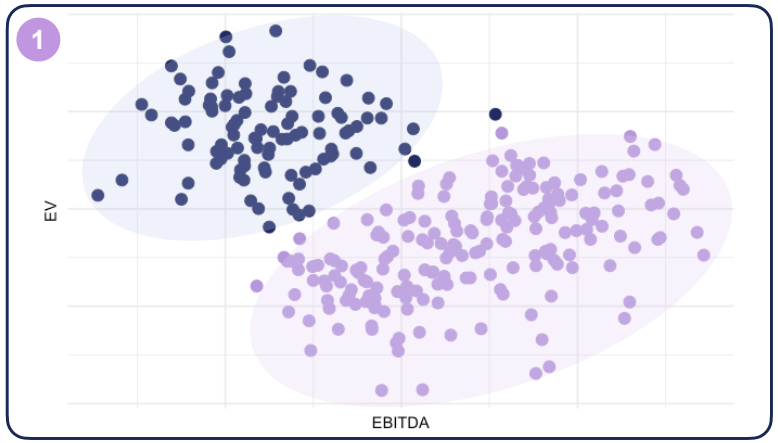
Allows setting the number of required clusters and dimensions

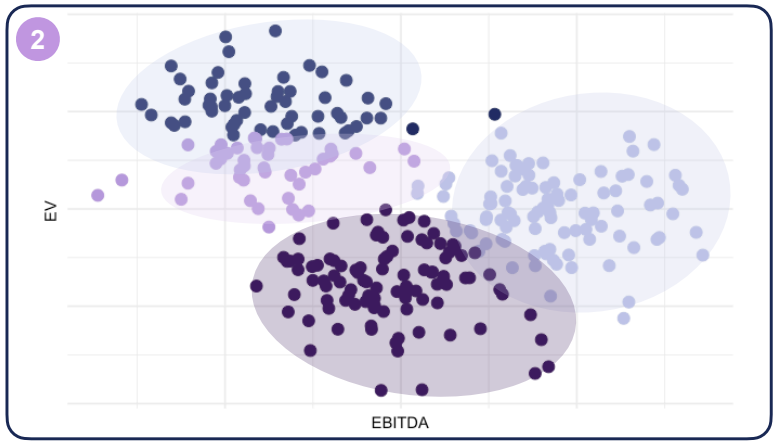
PEER GROUP MODEL BASED ON K-MEANS CLUSTERING

1. Clustering the data set based on an industry
2. Clustering the industry clusters based on assigned weights of variables
3. Formed clusters within an exact quarter
4. Final clustering with optimized weights

K-means clustering in two dimensions



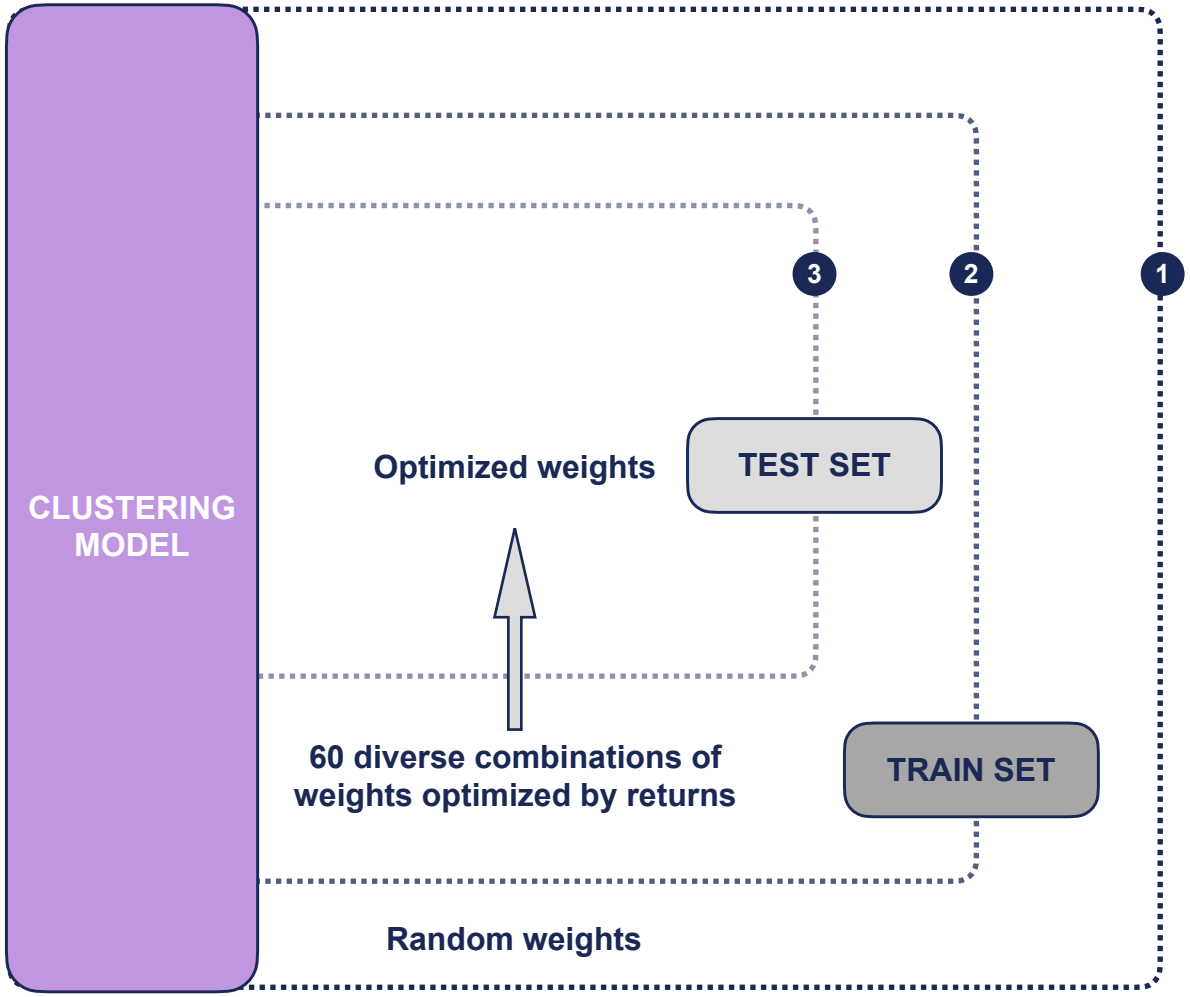




Optimization

Optimization by returns was performed with 60 weights combinations

Mechanic of optimization



Reasons for optimization

The clustering model finds peer groups based on weights assigned to **the input variables**

The assigned weights should be **chosen**

Automatic weights optimization generates weights that give **the highest return**

Steps of optimization

Calculate optimal weights of variables on the train set

Apply weights of variables on test set

Form peer group on test set

ChatGPT approach to peer group forming

ChatGPT 3.5 was used for forming the peer groups

W U T I S

What is ChatGPT and how to use it?

OpenAI

- AI-based chatbot
- **G**enerative **P**re-trained **T**ransformer
- Artificial neural network
- Transformer – deep-learning model – weighs input based on its significance
- Later models include more trainable parameters and training

Comparison of ChatGPT 4 versus ChatGPT 3.5 Turbo

GPT 4

- Larger model size: ~ 1T parameters
- Complex tasks
- Higher accuracy
- Extensive training dataset

GPT 3.5 Turbo

- Smaller model size: 175B parameters
- Lower computer power
- Faster computation

The procedure of finding a peer group with ChatGPT 3.5 Turbo

1. **Using the predefined test set (~1000 companies with data for 20 FQs each)**
2. **Finding all companies in the test set that belong to the same primary industry**
3. **Sending GPT an anonymous peer group request**
4. **Ensuring the output corresponds to the set format and conditions**

The format of the anonymous peer group request to Chat GPT

Company X operates in the industry and geographically in region.
In financial quarter, company X has the following financial data:
a beta of [ ], total assets of [ ], total revenue of [ ], an EBITDA margin of [ ],
a debt-to-assets of [ ], a cash-to-assets of [ ], and an EV of [ ].
The firm also has the following ratios: an EV/Sales of [ ], an EV/EBITDA of [ ],
and a P/E of [ ].
Given these values, pick between 8 and 12 companies from the following:
all test companies with the same primary industry, to **use** as a peer group for
Company X **in** the financial quarter given.

Strategy & Conclusion

The strategy based on different models led to different results

Strategy overview

PEER GROUP

EV | EBITDA | EBIT | Sales

PEER GROUP MEDIANS

Valuation object: EV, EBITDA, EBIT, Sales > PEER GROUP MEDIANS → SELL

PEER GROUP MEDIAN

Valuation object: EV, EBITDA, EBIT, Sales < PEER GROUP MEDIANS → BUY

Clustering & ChatGPT

CLUSTERING MODEL

Cumulative Total Return

206.4%

CHATGPT MODEL

Cumulative Total Return

20.2%

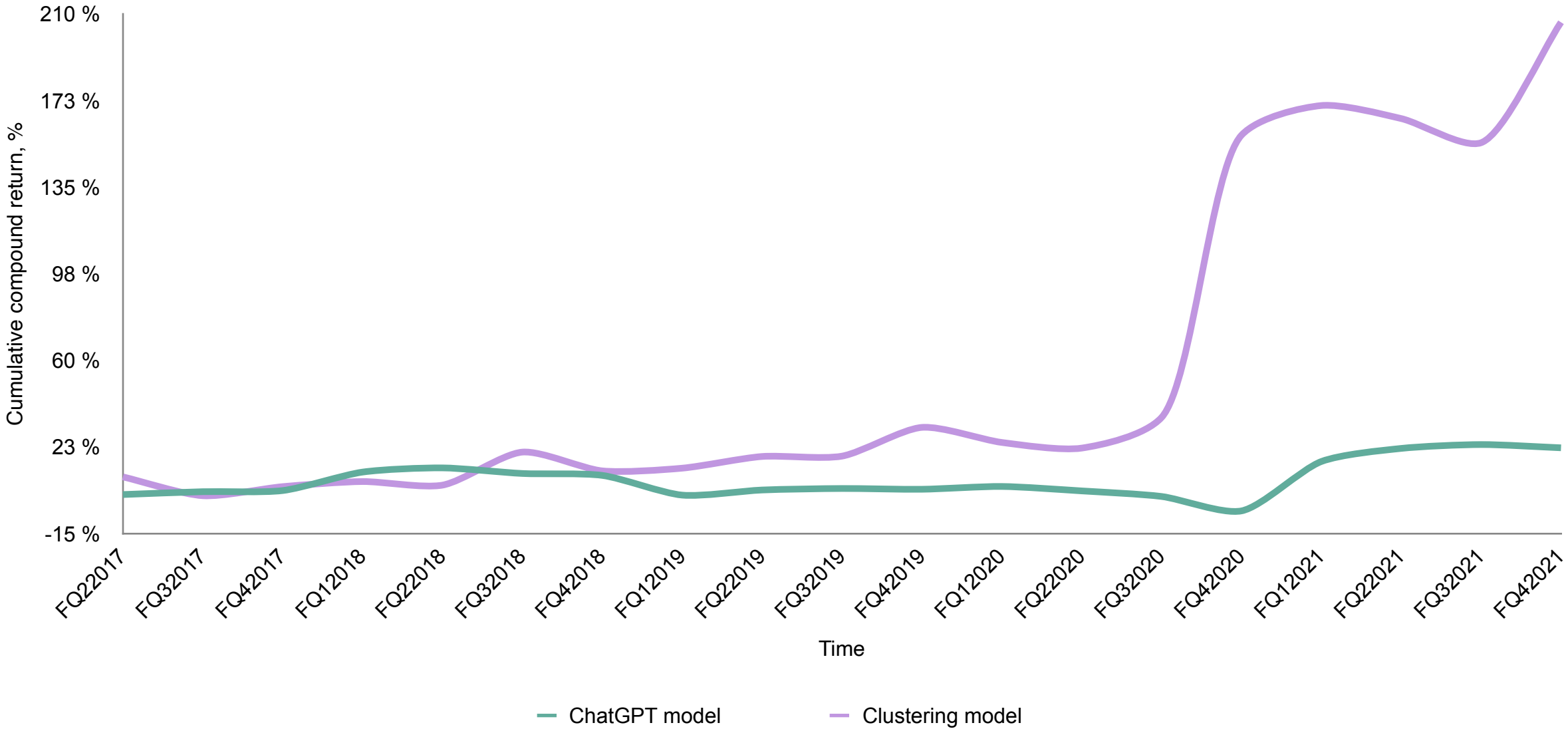
Improvements & Recommendations

Selecting companies for the test set by an analyst

Optimizing weights on a higher number of combinations

Clustering model & ChatGPT model

The clustering model generates cumulative compound return higher compared to return of ChatGPT model



Comparison to benchmark

The clustering model generates cumulative compound return higher compared to return of S&P 500