

Support of Investors' Decision Making in Economic Experiments Using Software Tools

Daria Kilinich^{1[0000-0001-6411-905X]} and Vitaliy Kobets^{1[0000-0002-4386-4103]}

¹Kherson State University, 27 Universitetska st., Kherson, 73003 Ukraine
vkobets@kse.org.ua, daria.kilinich@gmail.com

Abstract. During making decision the logit and probit patterns serve to resolve different issues based on statistical data regarding expediency or inappropriateness: opening LTD, investing funds, hiring employees, entering a new market, introducing innovations, etc. The purpose of the research is to support the decision making in economic experiments using software tools and logit and probit analysis. To achieve this goal, the following tasks are defined: investigation of the range of application of the logit and probit models; calculation of open data using the RStudio; development of decision support models using open data sources.

Methods and technologies of research: logit and probit models to predict the probability of dealing between traders of cryptocurrencies, cluster analysis of investor profiles through principal component analysis.

To distinguish different types of investors we can use cluster analysis which help us to reveal main types of risk-attitude investors. After that we can construct correspondence between specific users and financial instruments.

Keywords: decision making, economic experiments, software tools, cryptocurrencies, cluster analysis

1 Introduction

We studied criteria which affect prices of cryptocurrencies [1] and found out that combination of supply, mining difficulty, trading volume, and news reaction for each date can predict more than 70% of the price (we used Bitcoin for research). R.C. Philips and D. Gorse studied how to predict cryptocurrency prices bubbles using epidemic modeling and human reaction on social media [2].

Also, S. Colianni, S. Rosales, and M. Signorotti investigated cryptocurrencies algorithmic trading techniques based on Twitter sentiments analysis [3]. C. Lamon, E. Nielsen, and E. Redondo studied cryptocurrency price changes based on news and Reddit sentiments [4]. Kim YB et al. in 2016 did significant research about how users activities in communities affected prices of cryptocurrencies [5].

All researches we have mentioned above show that users activities affect prices. However, we applied a different approach in this research. Our idea was to predict cryptocurrencies prices based on their daily trading volume.

The decentralized digital currency Bitcoin presents an anonymous alternative to the centralized banking system and indeed enjoys widespread and increasing adoption

[6]. The digital currency market is considerably growing, especially in the most recent years. Level of uncertainty in returns has significantly increased during the high-price time period. The high-price regime phase has profoundly revealed consistent nonlinear dynamical patterns in the Bitcoin market [7]. The virtual currency supply is exogenous and therefore plays only a limited role in the price formation. Bitcoin is a digital currency based on a peer-to-peer payment system managed by an open source software and characterized by lower transaction costs, greater security and scalability than fiat money and no need of a central bank [8]. Bitcoin will remain a niche currency. Authors [9] analyze the time-varying behavior of long memory of returns on Bitcoin and volatility 2011 until 2017, using the Hurst exponent. R/S method is prone to detect long memory. Price volatility, measured as the logarithmic difference between intraday high and low prices exhibits long memory during all the period. This reflects a different underlying dynamic process generating the prices and volatility.

The creation of cryptocurrencies has changed FinTech industry and it continues to change it today, whereas people think that during 9 years nobody has found the real use of cases for blockchain technology [10]. Now people still depend on banks, because most countries did not define cryptocurrencies as national currencies; but in the future the decentralized systems, such as Bitcoin, can substitute traditional currencies. Also, due to continuously increasing digital society, financial services providers are looking to offer their customers the same services to which they are accustomed but in a more efficient, secure and cost-effective way.

In addition to mining (the process of extraction of the cryptocurrency), trading with cryptocurrencies is popular nowadays. It is risky but on the other hand, it is a fast way to get a great sum of money. For example, at the beginning of 2017, Bitcoin cost lower than \$1000 but in December 2017 it cost almost \$20000.

The **purpose** of the paper is to support the investors' decision making in economic experiments using software tools.

The paper is organized as follows: chapter 2 characterizes logit and probit models for data analysis; chapter 3 includes analysis of cryptocurrency data for trading; chapter 4 describes cluster analysis for investors' profiles that plan to invest in cryptocurrencies; the last part concludes.

2 Logit and Probit Models for Data Analysis

Logit model is a regression model, where a dependent variable can have only two alternative values "0" and "1". If dependent variable has more than two alternative results can be analyzed in a multi-vector logistic regression. In economic sense logistic regression is an example of a qualitative response to a discrete choice of decision maker. The probability of an event is determined by the function (1):

$$p_i = F(Z_i) = \frac{1}{1 + e^{-Z_i}} \quad (1)$$

where Z is a linear combination of independent factors.

The probit model is most often evaluated by probit regression using the maximum-likelihood method. Assume that the response variable Y is binary, that is, it can have

only two possible results, which we will denote as 1 and 0. We also have a regression vector X , which affects the result Y , then the model takes the following form (2):

$$\Pr(Y = 1|X) = \Phi(X^T \beta), \quad (2)$$

where \Pr – probability; Φ – cumulative distributive function of standard normal distribution; β – parameters of maximum-likelihood estimation. In the matrix form the regression will take following form:

$$Y^* = X^T \beta + \varepsilon, \quad (3)$$

where $\varepsilon \sim N(0, 1)$. Then Y can be considered as expression (4) if hidden variable is positive:

$$Y = \begin{cases} 1 & Y^* > 0 \\ 0 & \text{other} \end{cases} = \begin{cases} 1 & -\varepsilon < X^T \beta \\ 0 & \text{other} \end{cases} \quad (4)$$

The use of a standard normal distribution does not lead to loss of universality versus the use of arbitrary standard and average deviation, since the addition of a fixed amount to the average can be offset by subtracting the same amount. The equivalence of the two models has following form:

$$\Pr(Y = 1|X) = \Pr(Y^* > 0) = \Pr(X^T \beta + \varepsilon > 0) = \Pr(\varepsilon < X^T \beta) = \Phi(X^T \beta) \quad (5)$$

Because the logit [11] and probit models [12] are very similar to each other, the algorithm for constructing them is the same:

1. Determination of the dependent variable and factors
2. Construction of an independent variable, as a linear combination of independent variables
3. Specification of the equation for the desired probability of an event
4. Conducting calculations (maximum-likelihood method)
5. Interpretation of results and evaluation of quality assessment

The differences between the logit and the sample of models are in the specification of the random component ε_i , namely [13]:

1. in the probit of the model $\varepsilon_i \sim N(0,1)$ (standard random variables with mathematical expectation 0 and dispersion 1)
2. in the logit of the model $\varepsilon_i \sim \text{logistic}$, $f(t) = e^{-t} / (1 + e^{-t})^2$ (special logistical distribution $N(0, 1.6^2)$ with mathematical expectation 0 and dispersion 1.6^2)

Let's consider how price of cryptocurrencies impacts on decisions of potential clients to buy ('1') or not to buy ('0').

3 Analysis of Cryptocurrency Data for Trading

Investigation of the dependence of the purchase and sale of cryptocurrency. Data for the calculation was obtained from a public site, <https://finance.yahoo.com/cryptocurrencies> (fig. 1).

It has been selected 5 the most popular cryptocurrencies such as: BTC (Bitcoin), ETH (Ethereum), BCH (Bitcoin Cash), LTC (Litecoin), NEO.

Entering as a dependent alternating $Y_{i(1-5)}$, equal to 1 to indicate that the currency has been purchased and 0 that shows that the cryptographic currency has not been purchased [14]. Introduced explaining variables in order to show the course of this or another cryptocurrency from the first November 2017 till the 30th April 2018 (fig. 2): X_1 - BTC, X_2 - ETH, X_3 - BCH, X_4 - LTC, X_5 - NEO.

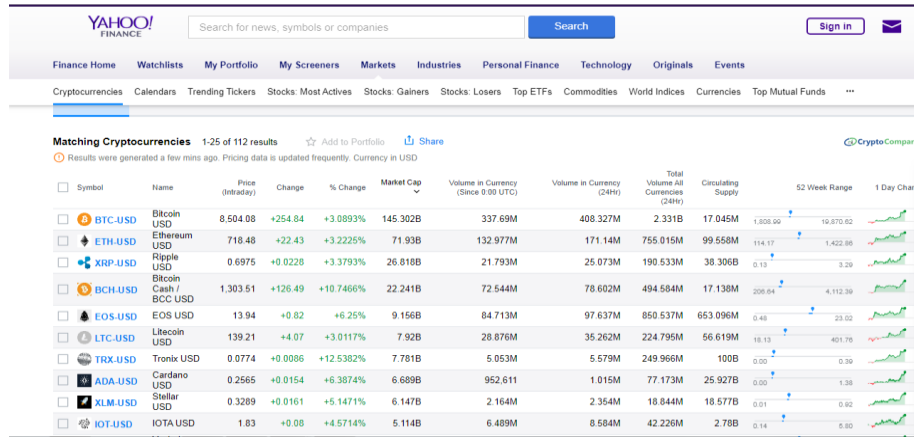


Fig. 1. Web-site of cryptocurrencies: <https://finance.yahoo.com/cryptocurrencies>

	A	B	C	D	E	F	G	H	I	J	K	L
	№	Дата	X1 (BTC-USD)	Y1	X2 (ETH-USD)	Y2	X3 (BCH-USD)	Y3	X4 (LTC-USD)	Y4	X5 (NEO-USD)	Y5
1	1	01.11.2017	6737,78	1	284,92	1	529,88	1	54,19	1	24,62	1
2	2	02.11.2017	7152,12	1	304,51	1	562,79	1	55,98	1	26,67	1
3	3	03.11.2017	7363,80	1	300,04	1	626,04	1	54,60	1	26,44	1
4	4	04.11.2017	7363,80	1	296,82	1	614,26	1	54,60	1	26,24	1
5	5	05.11.2017	7389,55	1	291,84	1	625,72	1	54,50	1	26,23	1
6	6	06.11.2017	6959,23	1	307,35	1	588,68	1	60,52	1	26,15	1
7	7	07.11.2017	7102,75	1	319,66	1	603,26	1	62,38	1	30,10	1
8	8	08.11.2017	7129,59	1	296,86	1	617,41	1	64,15	1	31,89	1
9	9	09.11.2017	7129,59	1	314,23	1	650,09	1	62,14	1	27,97	1
10	10	10.11.2017	6565,80	1	314,60	1	995,40	1	62,14	1	28,39	1
11	11	11.11.2017	6339,86	1	334,72	1	1325,56	1	58,54	1	26,82	1
12	12	12.11.2017	6522,45	1	334,72	1	1374,39	1	61,00	1	28,22	1
13	13	13.11.2017	6522,45	1	331,20	1	1346,96	1	62,13	1	29,59	1
14	14	14.11.2017	6597,06	1	330,32	1	1251,63	1	63,16	1	29,23	1
15	15	15.11.2017	7853,68	1	331,72	1	1187,03	1	70,70	1	28,74	1
16	16	16.11.2017	7853,68	1	346,65	1	896,51	1	67,36	1	40,20	1
17	17	17.11.2017	7699,95	1	354,60	1	1175,54	1	69,42	1	42,93	1
18	18	18.11.2017	8042,64	1	367,71	1	1243,86	1	72,38	1	40,00	1
19	19	19.11.2017	8042,64	1	360,52	1	1176,66	1	69,91	1	35,52	1
20	20	20.11.2017	8244,89	1	380,84	1	1245,28	1	69,91	1	34,37	1
21	21	21.11.2017	8099,97	1	406,57	1	1169,90	1	72,03	1	36,01	1
22	22	22.11.2017	8234,55	1	470,43	1	1298,62	1	72,94	1	34,67	1
23	23	23.11.2017	8013,41	1	470,54	1	1662,21	0	77,54	1	34,76	1
24	24	24.11.2017	8200,80	1	475,24	1	1625,05	0	88,79	1	34,76	1
25	25	25.11.2017	8200,80	1	466,27	1	1546,22	0	90,99	1	38,01	1

Fig. 2. Initial data of cryptocurrency rate

In order to track the dynamics of the cryptocurrency exchange rate at the initial stage of calculations, we have created a chart (fig. 3). With a help of it we can make out that only the BTC has significant fluctuations in value; the other four currencies have minor fluctuations.

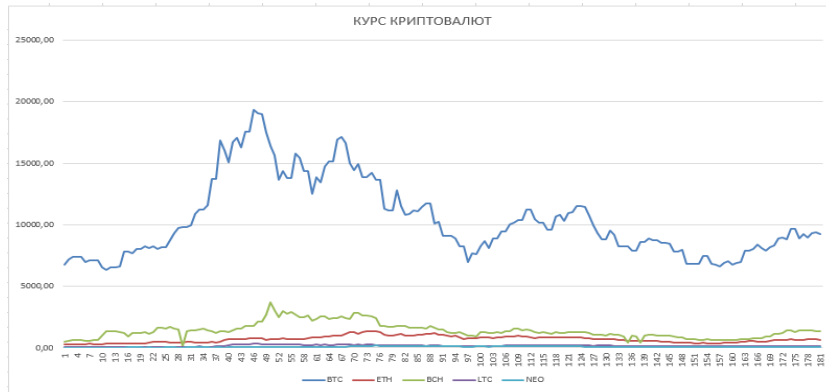


Fig. 3. Change of the cryptocurrencies rate for 181 days

So let's start work directly with RStudio. Download the data from MS Excel to RStudio (fig. 4):

```
Data <- read_xlsx("C:/RStudio/cr.xlsx")
View(Data)
```

	Дата	INDEPENDENT1	INDEPENDENT2	INDEPENDENT3	INDEPENDENT4	INDEPENDENT5	DEPENDENT
1	2017-11-01	6737.78	284.92	529.88	54.19	24.62	NA
2	2017-11-02	7152.12	304.51	562.79	55.98	26.67	NA
3	2017-11-03	7363.80	300.04	626.04	54.60	26.44	NA
4	2017-11-04	7363.80	296.82	614.26	54.60	26.24	NA
5	2017-11-05	7389.55	291.84	625.72	54.50	26.23	NA
6	2017-11-06	6959.23	307.35	588.68	60.52	26.15	NA
7	2017-11-07	7102.75	319.66	603.26	62.38	30.10	NA
8	2017-11-08	7129.59	296.86	617.41	64.15	31.89	NA
9	2017-11-09	7129.59	314.23	650.09	62.14	27.97	NA
10	2017-11-10	6565.80	314.60	995.40	62.14	28.39	NA
11	2017-11-11	6339.86	334.72	1325.56	58.54	26.82	NA
12	2017-11-12	6522.45	334.72	1374.39	61.00	28.22	NA
13	2017-11-13	6522.45	331.20	1346.96	62.13	29.59	NA
14	2017-11-14	6597.06	330.32	1251.63	63.16	29.23	NA
15	2017-11-15	7853.68	331.72	1187.03	70.70	28.74	NA
16	2017-11-16	7853.68	346.65	896.51	67.36	40.20	NA
17	2017-11-17	7699.95	354.60	1175.54	69.42	42.93	NA
18	2017-11-18	8042.64	367.71	1243.86	72.38	40.00	NA
19	2017-11-19	8042.64	360.52	1176.66	69.91	35.52	NA
20	2017-11-20	8244.89	380.84	1245.28	69.91	34.37	NA
21	2017-11-21	8099.97	406.57	1169.90	72.03	36.01	NA
22	2017-11-22	8234.55	470.43	1298.62	72.94	34.67	NA

Fig.4. Entering a table with data for calculations in the program RStudio

To start the calculation of regression and to work with it, we would enter the code of called 'mylogit' [15] (fig. 5) and output the result using the 'summary' function. It is clear due to this function that we want to predict the dependence of buying a cryptocurrency from the value of the rate on it. As an argument we specify: dependent and independent variables; the location of the initial data; 'family' indicates that the distribution type is binomial.

```

> summary(mylogit)

Call:
glm(formula = DEPENDENT ~ INDEPENDENT, family = binomial(link = "logit"),
    data = Data, na.action = na.pass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.934e-04 -2.000e-08  2.000e-08  2.000e-08  5.847e-04

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4004.700 266688.912   0.015   0.988
INDEPENDENT   -0.386   25.708  -0.015   0.988

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2.4155e+02  on 180  degrees of freedom
Residual deviance: 6.5132e-07  on 179  degrees of freedom
AIC: 4

Number of Fisher Scoring iterations: 25

```

Fig. 5. The calculation of the dependency of the dependent variable from the independent

A Result has been obtained, according to the results of the calculation that shows remainders and coefficients. Since the calculation of regression [16] is made, we are more interested in the coefficients from which the following picture is seen. The Pr (> | z |) indicator shows whether the coefficients are statistically significant or not.

Since in this case Pr = 0,988 it means that statistical significance exists. This calculation shows that the change in the rate of cryptography with a probability of 98.8% affects the decision to buy / sell cryptocurrency. The buyer during decision making compares and analyzes the cryptos and then he/she chooses cheaper cryptocurrency. So, the change in the price of cryptocurrency with a probability of 98.8% affects the decision to buy / sell cryptocurrency.

But the most important is the value -0,386, it means that in spite of increasing the cost of cryptocurrency of 1 currency unit the value of the logarithm decreases by 0,386 or 3,86%. But actually, these data form are not quite convenient to interpret, it's much more better to make the logarithm to the odds ratio through the exponent:

```

> exp(mylogit$coefficients)
(intercept) INDEPENDENT
Inf          0.679769

```

So, after calculating the exponent, we can say that with the increase of independent variables (price) for 1 currency unit, the ratio of chances of buying a cryptocurrency increases in 0.68 times. The next step of the calculation is to calculate the general level of significance (adequacy) of the model. This action can be done in following way: compare the residual deviation of the model with the deviation of the zero model; calculate the number of degrees of freedom; determine the level of significance correspondingly:

```

> mylogit$null.deviance-mylogit$deviance
[1]241.5509

```


The next step is to construct the 'True positive rate' curve (fig. 7), using previously obtained data.

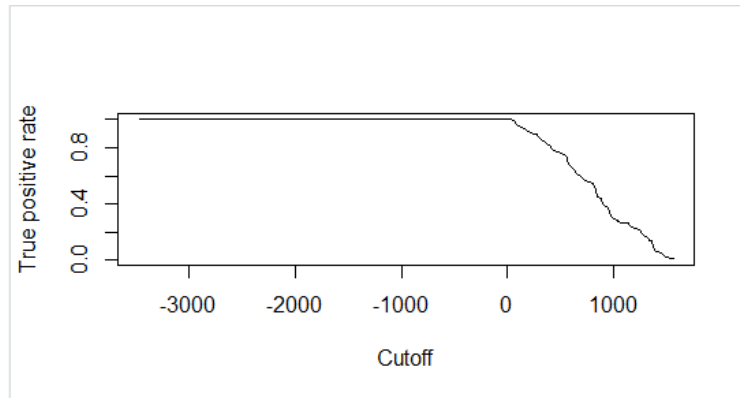


Fig. 7. Graph of the ROC curve

Let's proceed to the calculation of the area under the curve, cause it is more rational and better for future calculations. In order to do this, we would insert an additional change called 'perf1', and we construct a graph (fig. 8).

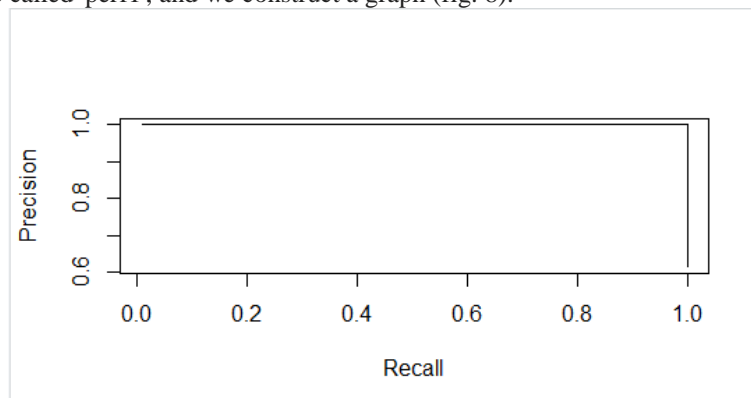


Fig. 8. Graph to calculate the area of the curve

In order to calculate the area under the curve of the graph, we need to use the function 'auc' (Area Under Curve).

```
> auc <- performance(pred, "auc")  
> auc <- unlist(slot(auc, "y.values"))
```

Having completed this calculation, the platform showed the 'auc' = 1 (> auc). It means that ratio of the number of correctly and incorrectly classified attributes to the selected values is perfectly suited.

Having completed the calculations, we have the following general form of the code (fig. 9), and the data (fig. 10).

```

1 Data<-read_xlsx("C:/RStudio/cr.xlsx")
2 View(Data)
3
4 mylogit<-glm(DEPENDENT~INDEPENDENT,Data,family=binomial(link = "logit"),na.action=na.pass)
5 summary(mylogit)
6
7 exp(mylogit$coefficients)
8
9 mylogit$null.deviance-mylogit$deviance
10
11 mylogit$df.null-mylogit$df.residual
12
13 dchisq(mylogit$null.deviance-mylogit$deviance, mylogit$df.null-mylogit$df.residual)
14
15 library(ROCR)
16 a<-predict(mylogit)
17 pred<-prediction(a,Data$DEPENDENT)
18
19 perf<-performance(pred, "tpr")
20 plot(perf)
21
22 perf1<-performance(pred, "prec", "rec")
23 plot(perf1)
24 auc<-performance(pred, "auc")
25 auc<-unlist(slot(auc, "y.values"))
26
27 minauc<-min(round(auc,digits = 2))
28 maxauc<-max(round(auc,digits = 2))
29 minauc<-paste(c("min(AUC)="),minauc,sep = "")
30 minauc<-paste(c("max(AUC)="),maxauc,sep = "")
31 legend(0.2,0.8,c(minauc,maxauc,"\\n"),border = "white",cex = 1.7, box.col = "red")

```

Fig. 9. Completed code

Data	
cr	181 obs. of 3 variables
Data	181 obs. of 3 variables
mylogit	List of 30
perf1	Formal class performance
pred	Formal class prediction
perf	Formal class performance
Values	
a	Named num [1:181] 1404 1244 1162 1162 1152 ...
auc	1
maxauc	1
minauc	"max(AUC)=1"

Fig. 10. Used data during calculations

4 Experiment Evaluation of Investors' Decision

Estimation of probability to be purchased or not for different cryptocurrencies gives us opportunity to develop investment plans [17] for investors with different investment goals and risk attitudes using open dataset (http://www.di.uniba.it/swap/financialrs_data_uniba.zip) of investors profiles.

In this case we quantified our ordinal data:

- Risk profile=[very low; low; normal; high; very high]=[1, 2, 3, 4, 5];
- Investment goals=[very low; low; normal; high; very high]=[1, 2, 3, 4, 5],
- Sex=[male, female]=[0, 1].

R tools can process our dataset using principal component analysis (fig. 11) to disclosure main types of investors to prepare investment plans for them using financial instruments such as cryptocurrencies [18, 19].

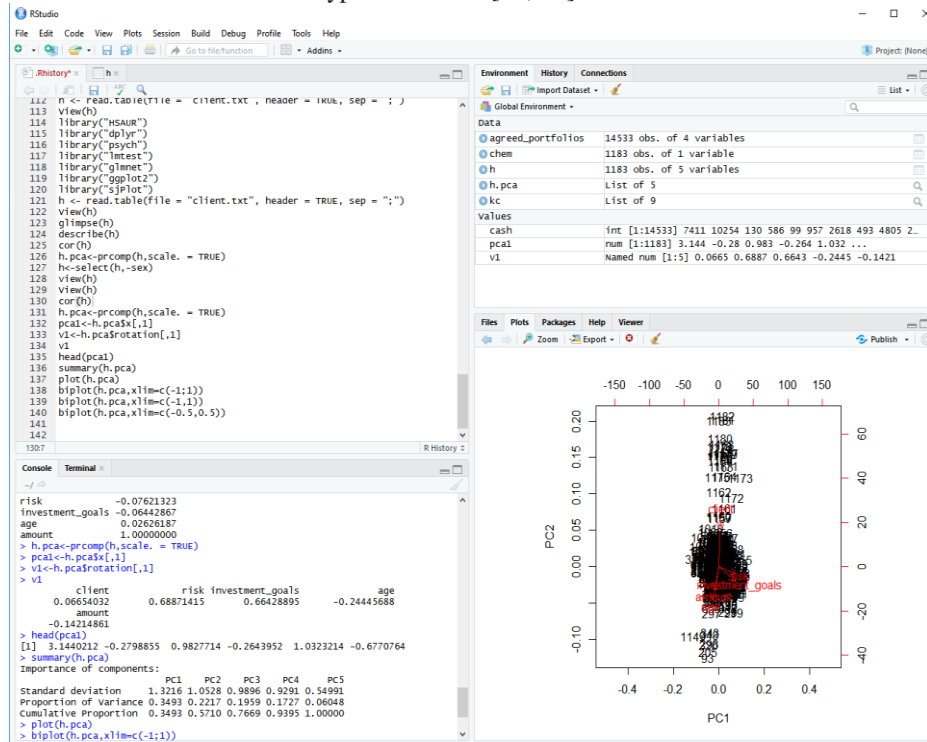


Fig. 11. Visualization of clusters for investors with different attitude to risk

Cluster analysis of estimated data for 14532 investors (fig. 12) revealed 3 types of investors:

- 1st type of investor: for the risk-aversing client, who invests in cryptocurrencies, the yield and the risk will be lower.
- 2nd type of investor: for the risk-seeking investor, the yield and the risk will be higher.
- 3rd type of investor: for neutral type of investor, the yield and risk will be lowest.

Principal component analysis using command biplot reveals that 1st main component includes risk (abscissa axis), whereas 2nd main component consists of investment goal (ordinate axis). The most investors are risk neutral, second largest group of investors (upper) is risk-averse. The shortest group (below) includes risk-seeking investors.

Thus investors who take part in trading of cryptocurrencies can be potential clients of financial services which construct different investment plans for different risk attitude clients and their behavior after price changing of cryptocurrencies.

Harmonization and Knowledge Transfer, Kyiv, Ukraine, May 14-17 2018, Vol.2105, 453-456 pp.

2. Phillips, R.C., Gorse, D.: Predicting cryptocurrency price bubbles using social media data and epidemic modelling, 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, 2017, pp. 1--7.
3. Colianni, S., Rosales, S.M., Signorotti, M.: Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis (2015). Mode of access: http://cs229.stanford.edu/proj2015/029_report.pdf
4. Lamon, C., Nielsen E., Redondo, E.: Cryptocurrency Price Prediction Using News and Social Media Sentiment. SMU Data Science Review 1(3), 1--22 (2017).
5. Kim, Y.B., Kim J.G., Kim W., Im J.H., Kim T.H., Kang S.J., et al.: Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. PLoS ONE 11(8): e0161197 (2016). <https://doi.org/10.1371/journal.pone.0161197>.
6. Ziegeldorf, J.H., Matzutt, R., Henze, M.: Secure and anonymous decentralized Bitcoin mixing. Future generation computer systems-the international journal of escience 80, 448--466 (2018).
7. Lahmiri, S., Bekiros, S.: Chaos, randomness and multi-fractality in Bitcoin market. Chaos solitons & fractals 106, 28--34 (2018).
8. Giungato, P., Rana, R., Tarabella, A. et al: Current Trends in Sustainability of Bitcoins and Related Blockchain Technology. Sustainability 9(12), 1--11 (2017).
9. Bariviera, A. F.: The inefficiency of Bitcoin revisited: A dynamic approach. Economics letters 161, 1--4 (2017).
10. Magnus, Y.R., Kagyshev, P.K., Perezetsky, A.A.: Econometrics. Initial Course: Case. (2007).
11. Ayvazyan S.A.: Applied statistics. Fundamentals of econometrics. Unity-Dana (2001).
12. Kaplan R., Urwitz G.: Statistical models of bond ratings: A methodological inquiry. Journal of Business 52(2), 231--261 (1979).
13. Efimova M.: Workshop on general theory of statistics: tutorial Bachelor's Guide: Yuright (2013).
14. Hosmer, D., Lemeshow, S.: Applied Logistic Regression (2nd Edition). John Wiley & Sons, Inc. (2000).
15. Moherny V. et al. Economic Encyclopedia: Publishing Center "Academy" (2008).
16. Dowherty K. Introduction to econometrics. INFRA-M (2009).
17. Kobets, V., Poltoratskiy, M. (2016): Using an Evolutionary Algorithm to Improve Investment Strategies for Industries in an Economic System CEUR Workshop Proceedings, vol. 1614, P. 485-501 (Indexed by: Sci Verse Scopus, DBLP, Google Scholar). <http://CEUR-WS.org/Vol-1614/ICTERI-2016-CEUR-WS-Volume.pdf>.
18. Ivanon, O., Snihovyi, O., Kobets, V. (2018): Implementation of Robo-Advisors Tools for Different Risk Attitude Investment Decisions, CEUR Workshop Proceedings, vol. 2104, P. 195-206.
19. Kobets, V., Yatsenko, V., Mazur, A., Zubrii, M. (2018): Data Analysis of Private Investment Decision Making Using Tools of Robo-Advisers in Long-Run Period, CEUR Work-shop Proceedings, vol. 2104, P. 144-159. <http://CEUR-WS.org/Vol-2104/>