

# Machine Learning II

## Module 5, 2019-2020

**Ilya Munerman**  
**Munerman & Partners, Interfax LAB**  
ivm@munerman.ru

### Course description

---

This course describes how dramatically changes of the information market provides new power tools for financial data processing and analysis. In addition, we will compare classical and contemporary tools performance. The main aim of this course is a species of different country data sets and technics for integration this data for common international data environment.

The course is designed for listeners which known elementary economics, finances, IT and mathematics and may be able for economists, IT specialists, managers, include MBA and journalists.

### Course requirements, grading, and attendance policies

---

Statistics, mathematics, corporate finance, assets valuation. The course grade is based on four home assignments (20%), case discussions (30%), and final exam (50%).

### Course contents

---

1. Contemporary financial analyses main challenges. Ontology and its role in the Contemporary financial analyses. Main categories: value, sustainability, dynamic. Understanding a difference between pattern recognition and prediction technics. Data quality and cognitive biases. Expert's professional crisis caused by data revolution. **Neuromorphic computing vs cognitive biases.**
2. Big, data mining, data science. Why general universe processing most powerful, then sample analysis? What is meaning sample today? Data source survey. Big data infrastructure. Low and ethical problem using big data. Main data types: open/close, government/private/corporate, free/paid, numerical/semantic. Data gathering, storage and retrieval instruments. Cloud technologies. Data base management, API, BI, ERP, final processing systems. Providers survey. From state to startups. Blockchain technology.
3. Main data types. Quote data for technical and fundamental stock market analysis. Labor data. Real estate data. Procurement and Contracts electronic trading platforms data. **B2B suggestion service. Procurement efficiency analysis.** Legal and court information. **Court cases liabilities estimation algorithm and prediction.** Web-sales and traffic data. Providers, aggregators, efficient and inefficient solutions. Open data hubs: pro et contra. Different solution costs. **Credit scores and paydex. Transaction data and cash boxes on-line date.**
4. Correlation. Correlation vs causalities. Metrics minding. Linear and nonlinear dependencies. Copulas.
5. Data processing. Modeling. Recognition, prediction and other forms of data processing. From regression to a neural network. Data mining. Cluster analyses and dendrograms. Stochastic processes. Probability vs reliability. **Classical ML – regression and clusterization, supervised and**

unsupervised learning, fuzzy logic and fuzzy c-means, SVM, logistic regression, PCA and another dimension reduction methods, CART, naive Bayes, etc. Ensembles methods – bagging (include random forests), boosting and stacking. Reinforcement learning, genetic algorithm, reinforcement learning for scorings and decision making. Neural networks MLP, GRNN, CNN, RBF etc.

6. Estimation and model testing. Statistical tests and criterions: R2, ROC-curve, type I and type II errors, graphical analyses. Ex-post testing.

7. Applications. Scorings, rankings, ratings. Ordinal and cardinal utilities. Tools for modern institutes: crowdsourcing, start-ups, franchising and so on. Credit conveyors. Value of credit and collateral value. Credit value monitoring according to regulation and standards. Real estate apps and valuation. Ad valor tax rates. Real estate databases. State and corporate real estate data hubs. Collecting data about all types of procurement: government, commercial, international, planned purchases. Bankruptcy analysis. News proceedings. Data standards problem. Auction houses data. AML and SAR, transaction fraud. **Fraud, failure and delinquency scores. Bankruptcy prediction without corporate finance.** Open societies, economic forecast, parameters estimation applications and so on. Visualization, infographics and data journalism. Importance, examples, tools, and applications. What is to be done. Contests and hackathons for provide original salvation. **Application for legal tech, prop tech, reg tech, ag tech, etc.**

## **Course materials**

### **Required textbooks and materials**

1. James G., Witten D., Hastie T., Tibshiriani R. (2015) An introduction to statistical learning with applications in R, 6th edition, Springer
2. 2. Brooks C. (2014) Introductory Econometrics for Finance, Third Edition, Cambridge University Press.
3. <https://ru.coursera.org/learn/machine-learning>

### **Additional materials**

1. D&B World bankruptcy reports
2. Certified fraud examiners exam questions
3. Special course materials will be provided in cloud storage

## **Academic integrity policy**

Cheating, plagiarism, and any other violations of academic ethics at NES are not tolerated.