

Вебинар «Об обучении в Магдебургском университете в рамках DAAD-проекта»

**Владислав Капустин
Дмитрий Легушев
(магистратура, КМиСА)**

Otto von Guericke University Magdeburg (OvGU)

- основан в 1993 г.
- 9 факультетов
- ~14000 студентов
- ~11700 публикаций в международные журналы
- академический персонал ~1770
- административный персонал ~1009

О совместном проекте БГУ и OvGU

- существует с 2005 г.
- курс по выбору основанные на материалах дисциплины читаемой на факультете математики OvGU “*Введение в оптимизацию*” от профессора Фолькера Кайбеля с последующей сдачей экзамена
- недельные поездки для студентов
- летняя практика
- обучение в магистратуре OvGU
- визиты немецких профессоров
- обучение на курсах немецкого языка на базе механико-математического факультета с преподавателем из DAAD

Учеба

- можно брать дисциплины с факультета математики и с факультета информатики
- возможность бесплатное посещение языковых курсов по немецкому языку
- за каждый экзамен дается 6 или 9 кредитов (в зависимости от количества пар в неделю)
- за семестр надо набрать 30 кредитов
- семинары
- нет учета посещаемости
- оценка на экзамене (как правило) не зависит от успеваемости в течение семестра
- на некоторых дисциплинах проводят промежуточные экзамены которые влияют на допуск к экзамену
- допуск к экзамену помимо этого может зависеть от процента сданных заданий
- отсутствие обязательных дисциплин

Факультет математики

- язык дисциплин: в основном немецкий
- домашнее задание сдавать не нужно, но нужно делать:)
- экзамены в основном устные
- институты
 - институт математической оптимизации
 - институт стохастики
 - институт алгебры и геометрии
 - институт анализа и численных методов

Примеры дисциплин

Vorlesungszyklus **Volker Kaibel**:

	Wann	Titel	SWS	Zielgruppe
K0	WS	Einführung in die Optimierung	4+2	B3
K1	WS	Kombinatorische Optimierung	4+2	B5, M1
K2	SS	Ganzzahlige Optimierung	3+1	B4,B6,M2
K3	WS	Geometrische Methoden der D.O.	4+2	M1
K4	SS	Algebraische Methoden der D.O.	3+1	M2

Vorlesungszyklus **Sebastian Sager**:

	Wann	Titel	SWS	Zielgruppe
S0	WS	Einführung in die Optimierung	4+2	B3
S1	WS	Nichtlineare Optimierung	4+2	B5, M1
S2	SS	G.-g. nichtlineare Optimierung	3+1	B4,B6,M2
S3	WS	Optimization Methods for ML	4+2	M1
S4	SS	Algorithmic Dynamic Optimization	3+1	M2

Nichtlinear Optimierung

Inhalt

1. Grundlegende Begriffe und Ideen der Nichtlinearen Optimierung
2. Unbeschränkte Optimierungsprobleme
3. Abstiegsverfahren, Konvergenzanalyse und Globalisierung

4. Newton-, Quasi-Newton- und Newton-Typ-Verfahren
5. Beschränkte Optimierungsprobleme
6. Verfahren der projizierten Gradienten
7. Aktive-Mengen-Verfahren zur Quadratischen Programmierung
8. Sequentielle Quadratische Programmierung (SQP-Verfahren)
9. Techniken zur Globalisierung von SQP-Verfahren
10. Barriereverfahren und Innere-Punkte-Verfahren
11. Software

Факультет информатики

- язык дисциплин: в основном английский
- нужно сдать определенный процент домашних заданий для получения допуска к экзамену (как правило $\frac{2}{3}$ или $\frac{1}{2}$)
- экзамены в основном письменные
- примеры дисциплин:
 - Swarm Intelligence,
 - Computational Intelligence in games,
 - Evolutionary Multi-Objective Optimization
 - Fuzzy logic, Bayes Networks
 - Advanced Topics in Machine Learning

Advanced Topics in Machine Learning

- для допуска к экзамену нужно было сдать $\frac{2}{3}$ домашних заданий и “защитить” проект
- по итогам курса сдается письменный экзамен

Overview (preliminary)



- Fundamentals of Data Science
 - Review of Machine Learning (Block Zero – 16th April)
 - Preprocessing (Block I – 23rd April)
 - Model Selection & Evaluation (Block II)
- Algorithms
 - Kernel machines and SVMs
 - Constrained Clustering
 - Semi-supervised learning
 - flat classification and clustering methods
 - hierarchical approaches
 - Markov Models
- Application related issues
 - Dealing with massive datasets
- Additional Material...

Advanced Topics in Machine Learning

1 of 5

Genre Identification on (a sub-set of) Gutenberg Corpus*

*Semester Project for Advanced Topics in Machine Learning

Uladzislau Kapustsin <i>Faculty of Mathematics</i> <i>Otto von Guericke University</i> Magdeburg, Germany uladzislau.kapustsin@st.ovgu.de	Dzmitry Lehusheu <i>Faculty of Mathematics</i> <i>Otto von Guericke University</i> Magdeburg, Germany dzmitry.lehusheu@st.ovgu.de	Vladimir Malyshko <i>Faculty of Mathematics</i> <i>Otto von Guericke University</i> Magdeburg, Germany vladimir.malyshko@st.ovgu.de	Daniil Sivachenko <i>Faculty of Computer Science</i> <i>Otto von Guericke University</i> Magdeburg, Germany daniil.sivachenko@st.ovgu.de
---	---	---	--

I. MOTIVATION

Feature extraction is very important process for data preparing, which can give better performance for machine learning models. It is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. In machine learning, pattern recognition and in text processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps. Feature extraction is related to dimensionality reduction. In terms of detection of genre of book on Gutenberg 19th Century English Fiction Book Corpus (selected 1k) data set we should consider the possibility for extracting as much relevant features as possible for improving results.

II. DATA SET

The original dataset consists of 996 books of 9 genres - Sea and Adventure, Literary, Love and Romance, Western Stories, Detective and Mystery, Christmas Stories, Ghost and Horror, Humorous and Wit and Satire, Allegories. Each book is represented by an HTML file with paragraphs separated by $< p >$ tags. After calculating the percentage of class instances for each class it became obvious that there's huge class imbalance (Fig 1). There were even classes with 2 or 3 instances, while other had more than 700.

Some instances from given dataset are meaningless. They are instances with IDs 465 and 622 which are just empty files without any text inside. And also ID 904 has too little information for any meaningful explanation of data. After the initial research we decided to make data splitting in order to balance somehow our classes (Fig 2).

III. CONCEPT

Concept

- We wanted to try out the following groups of features:
 - Features connected with the number of various parts of speech

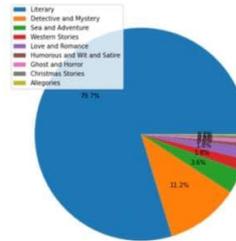


Fig. 1. Class balance in original dataset

- Features connected with text's sentiment
- Other statistical features connected with text's difficulty
- doc2vec embeddings

As methods for text classification we selected the following:

- Logistic regression
- Random forest
- SVC
- 1vsALL(SVC)
- Firstly according to the nature of the problem we definitely should apply feature extraction methods. We represent features using Sentiment analysis, Statistics to determine readability, complexity and grade level of a corpus and Part-of-Speech tagging method as well as

Evolutionary Multi-Objective Optimization

- Some preferences will not be considered as strongly as the DM prefers.

Example: Selecting a wife (here a maximization problem)

	Beauty	Cooking	housewifery	tidiness	Results
Mary	1	10	10	10	<u>6.4</u>
Jane	5	5	5	5	5
Carol	10	1	1	1	4.6
Weights	0.4	0.2	0.2	0.2	



Экзамены

- письменные
 - в основном информатика
 - зачастую очень похожи на то, что было на практических занятиях
 - на некоторые даже можно брать шпаргалку размером с А4 с различной информацией с лекций
 - нужно много считать и порой даже без калькулятора:)
- устные
 - в основном математика
 - очень дружелюбный? формат
 - можно самому выбирать что отвечать
 - никаких билетов (почти)
 - можно сдавать как на немецком, так и на английском

Дистанционное обучение

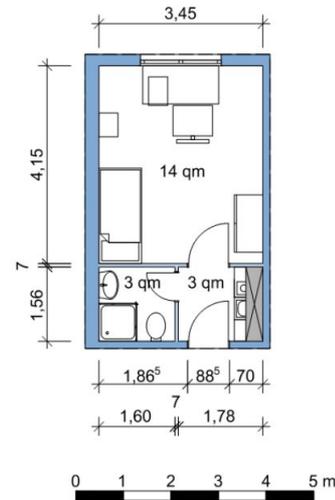
- система управления обучением moodle
- занятия как правило проходят в zoom
- по некоторым дисциплинам выкладывают записи лекций и слайды для асинхронного изучения материала
- онлайн конференции по практическим занятиям
- онлайн конференции для обсуждения материала лекций
- онлайн презентация проектов

Общежитие

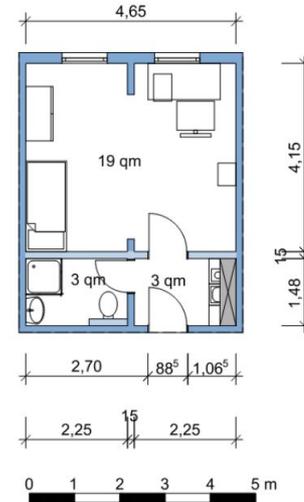


Общезитие

Grundrisse 1-Raum-Apartment One bedroom apartment plan



Typ 1, möbliert/ type 1, furnished



Typ 2, möbliert/ type 2, furnished

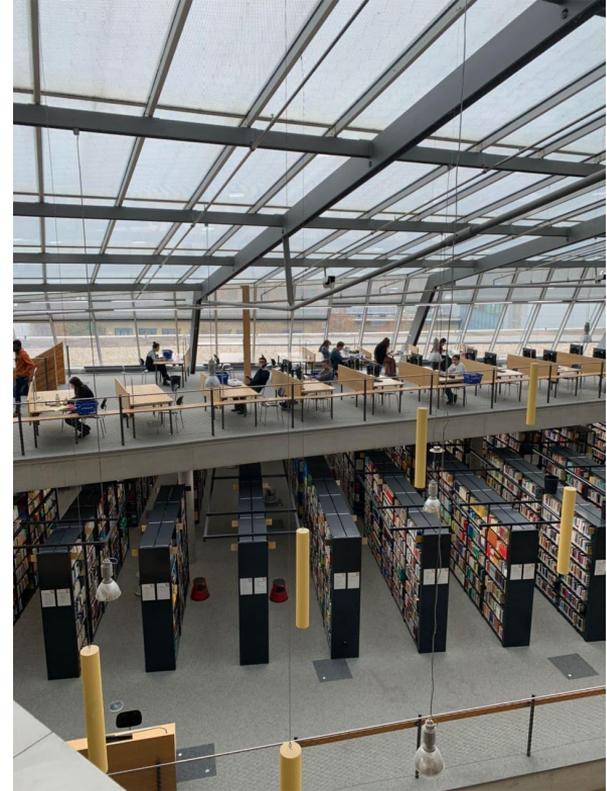
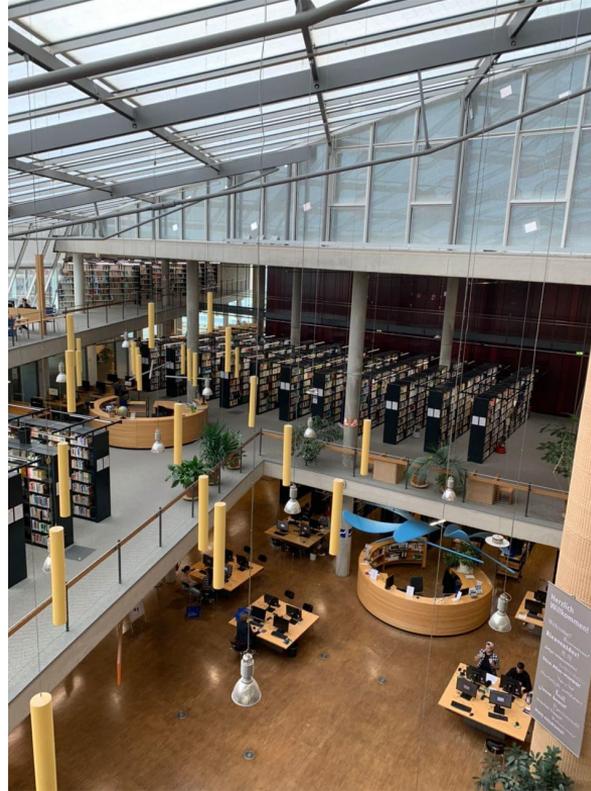
Общежитие



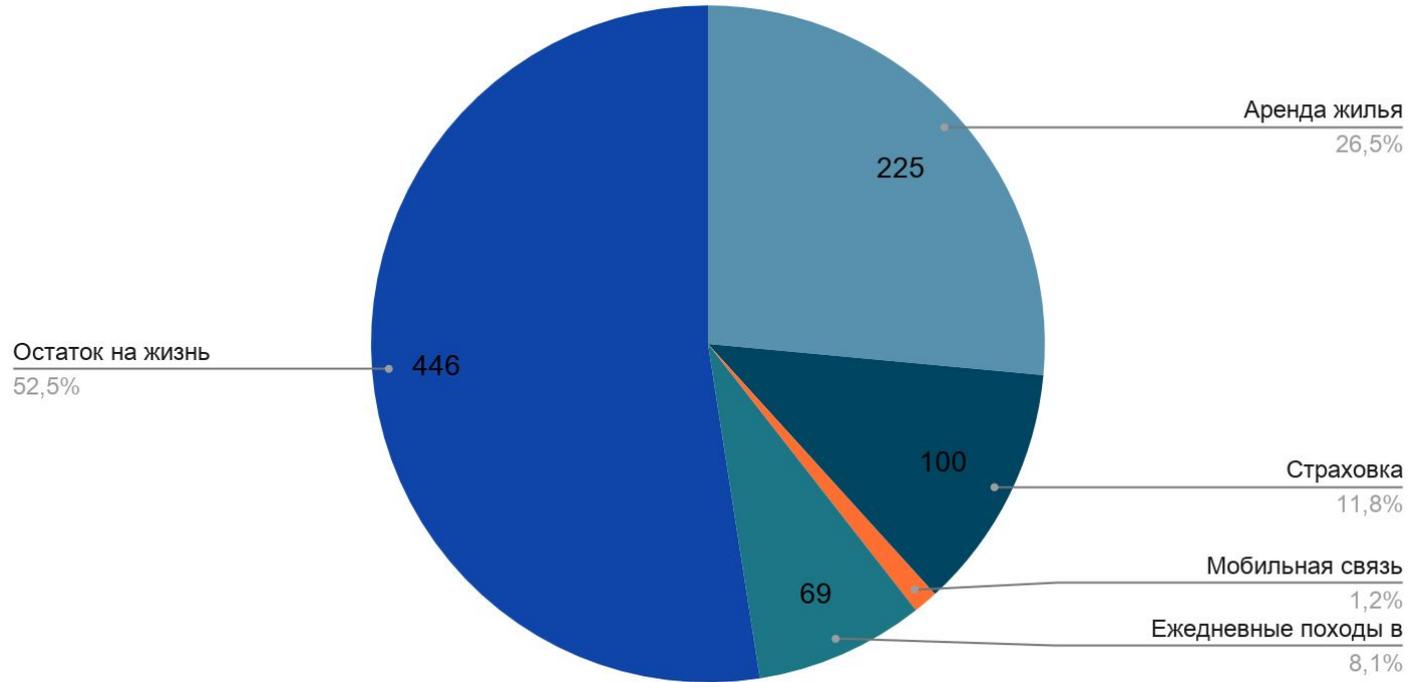
Кампус



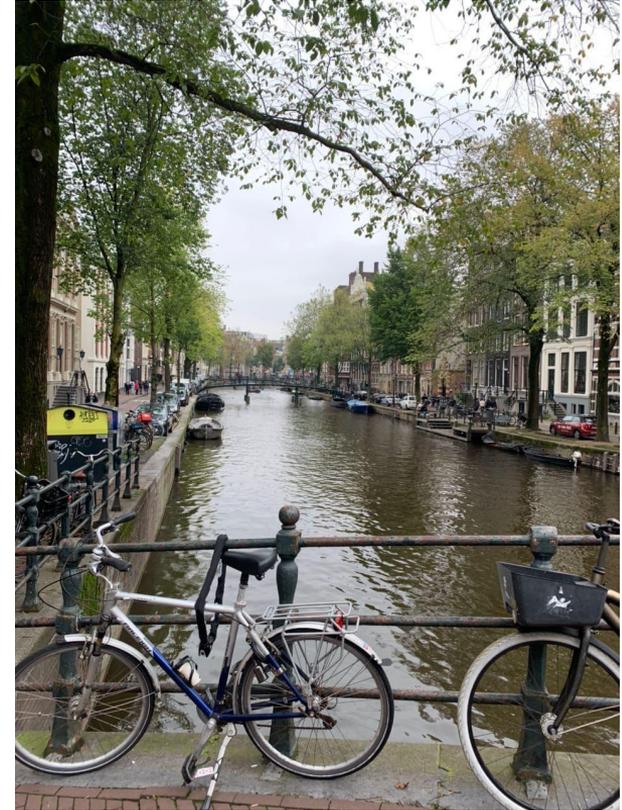
Библиотека



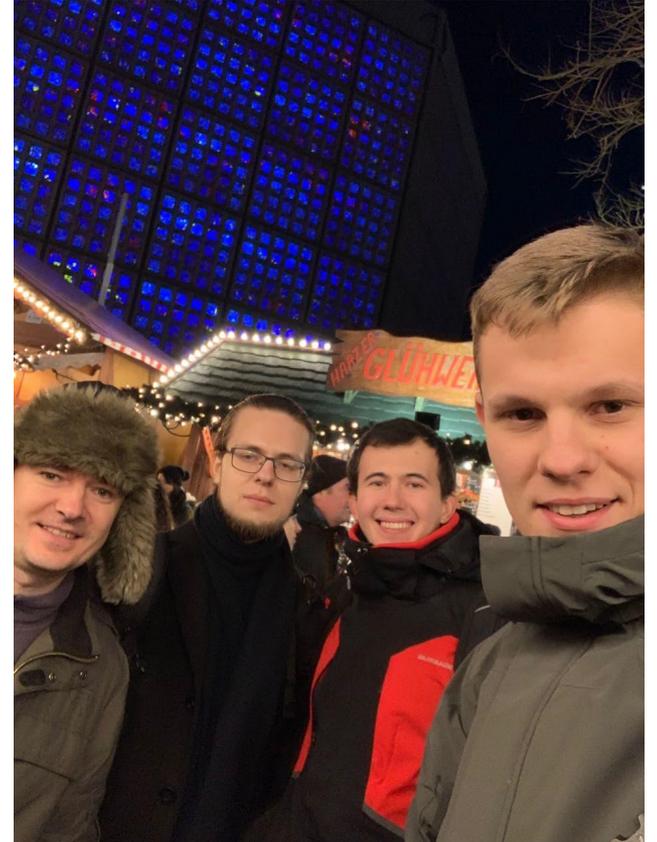
Финансы



Путешествия







Спасибо за внимание!

