Handelshøyskolen BI

MAN 51001 Analytics for Strategic Management

Term paper 60% - W

Predefinert informasjon

Startdato:	06-09-2022 09:00 CEST	**Termin:**	202310
Sluttdato:	04-05-2023 12:00 CEST	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	P		
Flowkode:	202310||10055||IN08||W||P		
Intern sensor:	(Anonymisert)		

Navn;

Anne Lenæs, Kenneth Myksvoll Means

Informasjon fra deltaker

Tittel *:	Machine Learning for prediction of grades.
Navn på veileder *:	Espen Andersen and Chandler Johnson

Inneholder besvarelsen konfidensielt materiale?:	Nei	**Kan besvarelsen offentliggjøres?:**	Ja

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	2
Andre medlemmer i gruppen:	

Term paper examination

Machine Learning for prediction of grades.

Hand-in date:

04.05.2023

Campus:

BI Oslo

Examination code and name:

MAN 51001 Analytics for Strategic Management

Programme:

Executive Master of Management

Content

Executive Summary

BI Norwegian Business School (BI), one of Europe's largest business schools, currently has an estimated budget of 125million NOK on expenses related to exams and grading in 2023. BI has set down a task force with a mission to cut these costs by 15 million NOK. Their report stated that further cost reductions would be possible with measures such as cut printing of digital exams, reduce exam elements for each course and increase the percentage each element has to the total assessment in each course and use digital home exams where possible to mention a few. Beyond this one might argue; the greatest potential for cost reduction lies in the use of AI as machine learning can be a powerful tool.

The project objective was initially to investigate whether a machine learning automation model could predict grades similar to a human grader. For several reasons, explained both under Academic Research and under Analysis we looked at what contributions it could give BI, namely a tool for predicting grades, but not replacing a human grader. Rather as a training tool for students. Our best performing model predicted with 90% accuracy if a submission was either (A, B, C) or (D, E, F).

Implementing an automation tool for grading, with the right refinements, can be cost-reducing, time saving and a quality consistency improvement opportunity.

Why should BI continue to investigate the opportunities that lie in Machine learning for predicting grades? Because the opportunities are nearly endless.

Our recommended first step in this direction is harmless, cheap, and useful in addition to allowing for incremental improvements. It could be the first step towards great savings.

Business objective and context

BI Norwegian Business School (BI) is an independent, not-for-profit foundation and the main provider of research-based knowledge on business and management disciplines in Norway. As one of Europe's largest business schools, BI seeks to attract, retain, and graduate students with a readiness to operate in an international context. Given the disruptive and drastic changes in Europe, a fast-changing financial situation for most organizations and an uncertain future, BI like many others, is forced to reduce costs and increase income.

BI invests substantial resources in exams and grading. Examples are salaries, rental costs, faculty work hours and more. BI also needs to prepare for changes in new legislation (Uhl) which will be implemented from August 1, 2023. Continuing today's practice without any changes, will increase grading costs by several million NOKs due to the requirements of two graders on all activities (e.g., class participations, class discussions, individual and group presentations etc.). Therefore, in November 2022, a BI Task Force was created with the objective of suggesting measures to reduce costs and make sure that BI adapts to the new act relating to universities and university colleges (UHL) taking effect in 2023. BIs current costs related to exams and grading is estimated to be 125 million NOK per year. The task force was asked to reduce this cost by at least 15 million NOK. Cost reduction should be even higher considering elements such as cut printing of digital exams, reduce exam elements for each course and increase the percentage each element has to the total assessment in each course and use digital home exams where possible to mention a few.

External factors affecting the global economy also have become a driver for reprioritizing budgets, and cost reductions on exams and grading have been set in motion. Expenses related to grading and exams are high, and moving from physical to digital exams is one cost reducing initiative.
Chatbots are considered enemies of higher education in terms of traditional exams and grading, as the sophistication of the tools has since early 2023 shown that the way exams are created needs to be changed. BI's top management has therefore asked all academic departments to address ChatGPT and exams. Given that digital exams are becoming the norm, automation of grading is a possibility and opportunity.

The project in brief

This project was always going to address machine learning automation models for predicting grades, but through the course of the project, the direction and focus of the project changed. In addition, it was nearly impossible to set direction for the project until we had data, as it was unclear what the data would contain, and therefore what problem the project would solve.

The initial project objective was to investigate whether a machine learning automation model could predict grades on text-based exams similar to a human grader. During the course of the project, for reasons explained under academic research and under analysis, we broadened the objective to look at what value automated grading of text-based exams could bring BI. Iterations with data exploration, data preparation, model training and model evaluation led us to conclude that much work remains before automated grading of text-based exams could replace human graders. Given the cost reduction goals and the potential with machine learning to reach those goals we have proposed a first step in this direction namely deploying a simplified binary classifier predicting if a text-based submission is either A,B,C or D,E,F as a training tool for students.

Research question

"How can automated grading of text-based exams create value for BI as an institution?"

"What value will this project provide?"

Exams and grading have always been a measurement of student learning, and in parallel students have worked on ways of how to fool / cheat the evaluation and developed advanced cheating skills. With introduction of ChatGPT and other similar tools, higher education institutions are faced with challenging innovation initiatives to find opportunities in AI and maintaining the quality of a BI diploma.

Digitalization and automation across industries forces disruptive actions. Higher education as an industry is changing. Covid-19 led to an acceleration of digital innovation and digital communication-tool adaptation. As a result, all industries went online, and for BI students,

expectations towards course delivery models, platforms and tools challenged traditional teaching.
This requires new ways of thinking about how to deliver courses, exams and grading. BI changed the standard for course delivery where 1/3 being asynchronous (own time, out of synch, pre-produced material) and 2/3 being synchronous (live, real-time) delivery. In addition, BI is now running a project on BI's future delivery model aiming to look at course delivery, usage of campus auditoriums, exams and grading to mention a few.

As described under the analysis chapter, how this project will add value to BI, changed from being a human labor-saving project with the intention of saving money while freeing up time for our faculty, to becoming a student learning project. Adding value to students by providing them with a tool for testing their knowledge and understanding of the course material. Specifically, their submission would receive a prediction (a number between 1 and 0) as to how likely it is that their submission is in the top half of the grading scale.

BI sees the long-term opportunities that lies in AI and ChatGPT both in teaching and learning, but in the short-term BI is addressing what can be done for the exams spring 2023, by encouraging knowledge and experience sharing on how to reformulate exam questions to ensure high quality assessments. AI will be able to support individual learning, and most organizations and industries are now asking themselves how to seize the opportunity that lies in AI. BI is currently engaging the Academic departments in sharing best practices for long-term policies related to AI. What potential and benefits of using machine learning might have related to exams and grading are still to be explored, especially on exams where there is not one right or wrong answer.

Recipients /Stakeholders of the analysis

BI Norwegian Business School is the main stakeholder, as this automation project has the intention of exploring what possibilities lie in using machine learning for predicting grades; a potentially cost-reducing, time saving and quality consistency improvement opportunity. In addition to the organization as a whole, faculty members, students and professional staff are also to be considered as both stakeholders and recipients of the solution presented in the analysis.

Faculty members as the main users of the grading-indication tool. Students, as they are the recipients of the potential value added in the form of indications as to how much of the material and content they understand. And finally, the professional staff who will oversee the implementation of and ensuring the quality of the system.

Internal analyses, industry research, and academic research

As one of the main providers of research-based knowledge on business and management disciplines in Norway, seeking ways of improving operations and competitive advantage, while at the same time finding solutions for foreseen and unforeseen changes, and disruptive innovations, is becoming more and more important. The phase of digital innovation, and sophistication of artificial intelligence tools such as ChatGPT, Google Bard, ChatSonic, and Claude to name a few, challenges business schools in more ways than expected only six months ago. Challenges can be exemplified with exams, where the Chatbot can write complete term papers, give well-written answers to exam questions, write analysis of cases and even predict a grade for an exam with quite impressive results.
For example, ChatGPT can be used for a variety of tasks in operations of an organization, amongst others as generating text content in a wide variety of types and styles, figuring our solutions by breaking down the core components of issues, automating responses for chatbots and helping developers with code by creating code, patterns and solutions for ML projects.

AI and machine learning are by far the biggest buzz words of 2023, and in most discussions at BI, the fear of what it threatens in operations shines through. New tech is challenging our core as knowledge and skills can now be taught and learnt everywhere. AI can be used to solve problems and help students cheat if we do not adapt our assessment of students. However, looking at one example where AI has revolutionized an industry is in chess. Stockfish, an open-source chess engine, trained on games played by humans. Alpha Zero, another chess engine, taught itself to play chess, and played like no human ever did. In return, humans learnt from it, and made chess even more interesting. This is a great example of how machine learning improves a sport, while also helping the population understand the game, and attracting an audience (Strogatz, (2018), Sadler, M., & Regan, N. (2019)).

Internal analysis

With the Covid-19 pandemic forcing an entire world into lock-down, BI made a decision to go online within 24 hours. At the same time BI was discussing how to deal with exams. All exams in the spring of 2020 became digital, and approximately 50 % of the exams needed to be changed to adapt to a digital home exam. This drastic change in assessment of students proved to be not only a feasible solution, but also a cost-effective solution. There were several challenges such as increased cheating amongst students, especially in some courses and disciplines. Nevertheless, this was the kick many institutions needed to explore new opportunities for exams that are maintaining the quality and reducing costs. In November 2022, with 4 semesters of digital exams, a task force was established at BI with the objective of suggesting measures to reduce costs with 15 million NOK. In parallel these measures are seeking to explore sustainable solutions for assessment.

Sustainable solutions for cost reduction are as mentioned earlier, cut printing of exams, including a copy of digital exams, reduce exam elements for each course and increase the percentage each exam-element has to the total assessment in each course and use digital home exams where possible to mention a few. Supporting the task force 's report, there are still cost reducing opportunities to explore; using machine learning is one such opportunity.

A new exam tool, and implications of ChatGPT

BI signed a contract with UNIwise in april 2019 after running exams through a self-developed Exam tool called DigiEx for more than 10 years. This was a decision following BIs digitalization strategy of purchasing solutions existing in the market rather than developing tools from scratch. DigiEx was running at maximum capacity with an estimated annual operating budget of 24 million NOK.

UNIwise is an international exam and assessment solution partner for higher education. The mission of UNIwise is to help educational institutions deliver open, transparent and reliable exams and assessment of the highest quality. Since 2012, UNIwise has developed and delivered WISEflow to educational institutions across Europe and today they have a customer portfolio covering 14 different countries. UNIwise is headquartered in Aarhus, Denmark with offices in Manchester, United Kingdom and Skien, Norway.

In October 2021, BI had its first digital exam using the new exam platform WISEflow. More than 600 students handed in their exam using their own device which was a first for BI. In 2022, 75000 papers were submitted in WISEflow. Thursday April 27, 2023, the main exam period at BI officially started. Some courses did start the exam period a little earlier, but for most exams, more precisely more than 88 500 exam papers will be submitted, over the next 8 weeks. Out of these 50% are home examinations, 50% are school examinations, but what is interesting for this project is that this semester more than 70% of BI's exams will be conducted digitally.

A platform or tool is not the solution by itself, as it is the content and way of formulating an exam that provides input for machine learning automation. In the case of BI, maintaining the current ways of creating exams will provide an opportunity for students to put the questions directly into ChatGPT and have this language model write the answers. Therefore, BI has sent out a request to all academic departments asking for their input when analyzing the implications and challenges of AI/ChatGPT vary across academic disciplines and departments. BI is dependent on the respective disciplines' insights, experiences, and best practices to provide useful policies. There might be very legitimate justifications for some applications of AI on assignments, but it is very difficult for an administrative unit to decide upon boundaries without disciplines' input.

Industry research

When looking globally at higher education institutions, no one has documented that they are AI driven in operations of course delivery and assessment. Most research is still at the stage of recognizing that AI will be an essential part of the future of Higher Education (Instituto Impresa, Imperial College). Abdous (2023), argues that AI is shaping the future of higher education as it has emerged as one of the most powerful agents of change, and presents the industry with unprecedented academic, ethical and legal challenges. AI has and still is quietly disrupted higher education's activities within teaching, leading, research and administration. When looking at potential solutions, the Higher Education industry is on a larger scale than before suggesting that interdisciplinary discussions on the implications and complexities AI will bring to the academic landscape. Connectedness and collaboration with industry and public sector seems to be a direction to move in implementing AI as an integrated, transparent and impartial partner for student learning. (Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A. et al, (2023)).

In the academic landscape, it seems that most higher education institutions are still exploring the opportunities that lie in AI, and the initiatives made are on an individual basis.
Asking what the main exam system providers are doing, the exam platform WISEflow, and UNIwise are at the moment just looking at this from a student perspective. In a partner meeting in April, they said: *"As with all new technology, AI text generators present both a threat and an opportunity for higher education. One the one hand – the positive view – if harnessed well, can ChatGPT be the catalyst to help realize the goal of more authentic assessment? On the other hand, some faculty may take the route of least short-term resistance and double down on detection, reference-checking and proctoring. Is the latter just a finger in the dyke?*
As a responsible provider and partner to HE, our position will evolve and develop as technology evolves and develops, and as we understand more about the consequences for HE. Guided by ethical principles and best practices, we will ensure that our platform supports universities' current and future assessment strategies in a sustainable way" (UNIwise, 2023).

What automation projects exist using grades and exams?

When searching for academic research on automation of grading in higher education, published after 2023, 6280 findings come up. Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A. et al. (2023) calls on researchers to take an active stance by developing better understandings of how digital automations can be meaningfully integrated into education and adopting a sense of realism (rather than idealism) when reflecting on where we might like to be going next which is supported by Gallagher and Breines (2022), and Gibson (2022).

Historically automation projects existed, and as early as in 1994, at the University of Maryland, Kassandra was launched. An automatic grading system for grading assignments in scientific computing. A student could interactively use this system to check the correctness of his program assignments. The grade for a correct solution was automatically recorded. As a result, they could ease the work of student assistants, by having the students themselves check their homework assignments on the computer, and correct solutions were graded automatically. The teaching assistants no longer needed to spend the majority of their time on the chore of grading assignments and could rather work on developing advanced test models, that were fed into Kassandra for the next cohorts to test and train on.

Another example is from NTNU. For a long time, they have had computer science students testing models for automation of grading. Already in 2018, students explored the opportunity of automating grading, but the results came back inconclusive. They argued that the dataset was suboptimal and ended up overfitting the classifier. Which illustrates that 2018 in this context is a long time ago.

Academic research

When diving into the literature and doing searches where data mining is used in higher education, most of the research found is where machine learning is used as a quality enhancement by identifying students at risk of low motivation or dropping out of their studies (Mahboob, K; Asif, R; Haider, N.G., (2023), Iatrellis, O., Savvas, I.K., Kameas, A. et al. (2020)). Machine learning becomes a supplement in detecting students that struggle and students that thrive, and complemented by other input data, the higher education institution has a low threshold for following up students, making them feel noticed and acknowledged. (Yağcı (2021), Baker & Yacef, 2009; cited in Fernandes et al., 2019)).
Until now, most research on predicting grades is conducted on either data science courses or on mathematical courses where there is a right and wrong answer. (ChatGPT, Mahboob, K; Asif, R; Haider, N.G., (2023), Iatrellis, O., Savvas, I.K., Kameas, A. et al. (2020)), but there is an increase in testing this on a wider portfolio of courses and academic programmes. One reason has been the COVID-19 pandemic, and its implications on traditional teaching methods. When using machine learning automation as a quality tool, higher education institutions can improve consistency through pattern recognition of student behavior and performance. This is of high value as higher education institutions are in transition from being run as traditional universities, to more demand-driven organizations.

Yağcı (2021), argues that educational data mining has become an effective tool for exploring the hidden relationships in educational data and predicting students' academic achievements. Educational data mining (EDM) is the use of traditional DM methods to solve problems related to education (Baker & Yacef, 2009; cited in Fernandes et al., 2019).

EDM has become an effective tool used to identify hidden patterns in educational data, predict academic achievement, and improve the learning/teaching environment. (Yağcı, 2021). In this study, the result showed that students' midterm exam grades are an important predictor to be used in predicting their final exam grades. With this knowledge, lecturers can use early recognition of students having a below average academic motivation, and they can match students with low academic motivation with children with high academic motivation and encourage groupwork in projects. Where this research has translational links to predicting grades is the aim of having influence on and improving student learning. By having the students testing their understanding of the course material by introducing automated grading on term paper drafts, the student instantly gets an indication of both their level of understanding and for instance using word clouds to provide more specific feedback as to what words or n-grams contributed to the predicted grade.

Automation is coming!

Digitalization of education is ongoing, and most higher education institutions have a high absorption and implement small automations into everyday operations and processes. (Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A. *et al, (2023),* Cerratto P. T, Lindberg, Y. & Buch, A (2023)). According to Kott (2022, cited in Cerratto P. T, Lindberg, Y. & Buch, A (2023)), technologies are quickly becoming woven into the digital infrastructure of education, and become 'invisible and forgettable', and will only be seen when they impact us unfairly, harmful or in other negative ways. Grading takes time, and setting labor-saving arguments aside, it is argued that automated technologies might be capable of performing better than a human teacher, as automated essay grading software can process student assignment at both an unimaginable scale and speed in comparison to a human grader. Consistency and fairness are also strengths talked about when automating grades Cerratto P. T, Lindberg, Y. & Buch, A (2023). Fears of automation are not far from the claims and fears in the 1950 on that robots will take over jobs, but looking at the history since the 1950's, what still shines through is the value of education and the role of pedagogical work, so in the academic voyage of a students, machine and human will for quite some time still go hand in hand. (c.f. Park and Humphry (2019) cited in Cerratto P. T, Lindberg, Y. & Buch, A (2023)).

Looking at how machine learning can encourage and support student learning by being an auto grading guidance tool predicting the likelihood of a grade on a submission, when asked; ChatGPT answers that automation of grading can benefit students in several ways.

"Overall, automation of grading can help provide students with more consistent, timely, and personalized feedback, improving the quality and fairness of the grading process and ultimately benefiting their learning outcomes, of which all are drivers for smooth operations in higher education".

When asked if automation of grading can replace humans, ChatGPT responded:

"*Automation of grading has been increasingly utilized in recent years, particularly in multiple-choice tests or standardized exams where answers can be easily compared against a pre-determined answer key. While automation of grading can save time and resources, there are limitations to its effectiveness and reliability. Automation of grading cannot fully replace humans for the following reasons: Subjectivity, error rates, lack of critical feedback and human touch. Ultimately, a combination of automated grading and human assessment can help to ensure accurate and fair evaluations while providing personalized feedback to students."*

Text is everywhere.

Most research on predicting grades is conducted on either Data Science courses or on mathematical courses where there is a right and wrong answer. While in contrast, in our everyday life, "text is everywhere", Provost and Fawcett explicitly claim (Chapter 10), and rightfully so. As a form of information, documentation and communication, text is the main "code" used. While text can be clarifying, it can also be difficult as text is often referred to as "unstructured" data. As text does not have the normal structure expected for data such as tables. In machine learning automation projects data rarely presents itself in a neat feature vector representation used as input in most data mining methods. Further description of the text and how it is used in this project can be found under Tools and techniques.

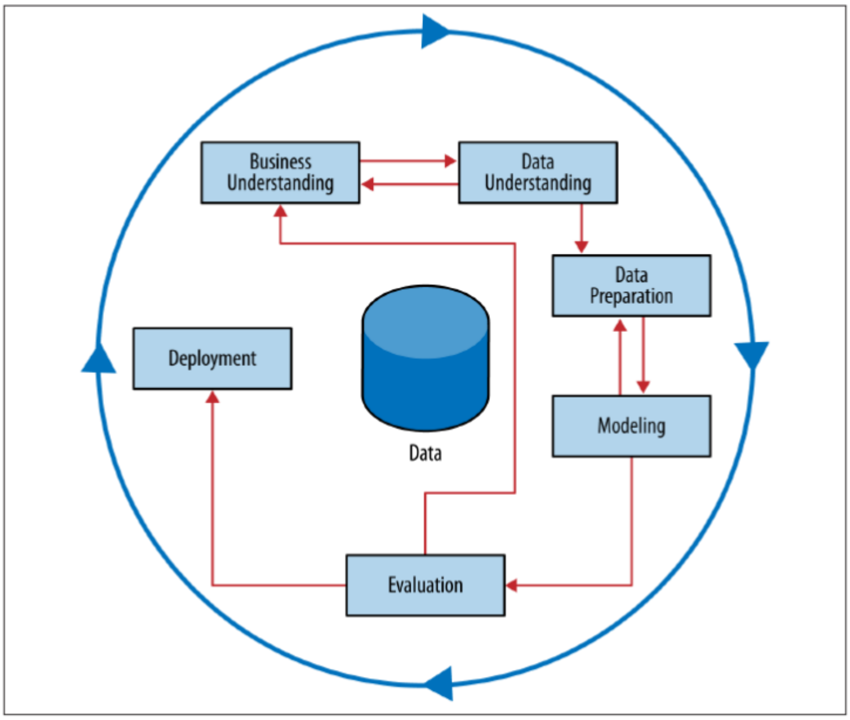
Internal analysis, industry research and academic research all point in the direction of ML automation of various activities is the way to go, however it can be seen as a challenge on how to get there. While this path is being created, machine learning automation projects for predicting grades and human assessment needs to go hand in hand as a combination to ensure accurate and fair evaluations while providing personalized feedback to students. (c.f. Park and Humphry (2019) cited in Cerratto P. T, Lindberg, Y. & Buch, A (2023).

Linking theory, business case and the analytics project together, the next part of this document will take on a more informal form, guiding the reader through the process, discoveries and iterations to overcome challenges towards a conclusion and further recommendations.

Analysis

Crisp-DM

We found it natural to work iteratively according to the Crisp-DM (Provost and Fawcett, 2013) methodology as illustrated by the diagram below. Discussing possible business value of a working model, exploring the data, preparing the data, training models, evaluating models and iterating back to either the data preparation phase or all the way back to discussions concerning where a model might add value to BI.



Crisp-DM figure (Provost and Fawcett, (2023, p. 27)

Data collection

For this project gathering data proved to be, as in most ML projects, quite challenging.
With Top Management approval, there were still several organizational approvals needed before the Digital department contacted UNIwise with the request to extract data.

This was the first time UNIwise extracted large amounts of data from the exam tool WISEflow, and they were not sure firstly on the process, secondly on the timeframe for accessing the data, and thirdly what the extracted data exactly would contain. The cost for extracting data was 400 EUROs. Under data description and data preparation the process of transforming the json files to tabular data is described in detail.

Tools and techniques

Python

Data preparation was done in Spyder 3.9. The python libraries used were pandas, html, re, requests, io, pypdf and sklearn. All the code used for data preparation is detailed in appendix B. A closer step by step description of the work done in python is also detailed under "Data description" & "Data preparation".

Auto ML

DataRobot was used for feature selection, model training, model evaluation and predictions. We did not second guess DataRobot's recommendations for model selection. Descriptions of the iterations from data preparation, model training, model evaluation gaining insights into potential business value of various models and back to data preparation revolved around the heavy lifting done by DataRobot. Work with evaluation of model performance metrics and scrutiny of confusion matrixes and word clouds produced by the various models is detailed below.

Crucial to understanding how it’s possible to make predictions based on text is TFIDF.

“An ML model can operationalize text using the Term Frequency-Inverse Document Frequency (TF-IDF) technique as a way to represent and quantify the importance of words in a collection of documents. TF-IDF is a feature extraction method that transforms raw text data into numerical vectors that can be used as input for machine learning algorithms.

Here's how the TF-IDF technique works:

1. *Tokenization: First, the raw text data is tokenized, meaning it's broken down into individual words or tokens. This process also often involves removing stop words (common words such as "the", "and", "is", etc.), punctuation, and converting all words to lowercase.*
2. *Term Frequency (TF): For each document, calculate the term frequency of each word. Term frequency is the number of times a word appears in a document divided by the total number of words in that document. This represents the local importance of a word within a document.*
3. *Inverse Document Frequency (IDF): For each unique word in the collection of documents, calculate the inverse document frequency. IDF is the logarithm of the total number of documents in the collection divided by the number of documents containing the word. This represents the global importance of a word across the entire collection of documents.*
4. *TF-IDF Calculation: Multiply the term frequency (TF) of each word in a document by its inverse document frequency (IDF). This results in the TF-IDF value for each word in the document. High TF-IDF values indicate that a word is important and relevant within a specific document and less common across the entire collection of documents.*
5. *Vectorization: After calculating the TF-IDF values for each word in a document, create a numerical vector representing the document. This vector has the same length as the number of unique words in the entire collection of documents, with each element corresponding to the TF-IDF value of a specific word. If a word does not appear in a document, its TF-IDF value in the corresponding vector is 0.*
6. *Input to ML Model: The resulting TF-IDF vectors for each document can now be used as input for a machine learning model. These numerical vectors can be fed into various ML algorithms, such as logistic regression, support vector machines, or neural networks, to perform tasks like classification, clustering, or regression on the text data.*

By using the TF-IDF technique, a machine learning model can effectively operationalize text data, transforming it into numerical vectors that capture the importance and relevance of words within a collection of documents."

Chat GPT

Chat GPT-4 proved to be an invaluable resource for us in completing this project. Particularly as an assistant in writing the necessary python code. We've included a few paragraphs in this paper where the bot is quoted directly answering technical questions outside of our expertise as demonstrated above.

Data description

The data initially asked for was exam submission text and grade for a selection of courses. Once we started unpacking the data, we discovered that we'd received more information than expected. The data provided by UNIwise was in a json format containing the following four columns:

- flowData dictionary with the following keys: flowTitle, subtitle, code & flowType.
- assignment nested dictionary with the main dictionary keys: assignment and appendices.
    - assignment sub dictionary had the keys: name, size, downloadUrl,
    - appendices sub dictionary was empty but would likely contain the same keys.
- assessors list of dictionaries where each dictionary in the list represents an individual assessor with the keys: assessorId, userId and name (actual assessor name).
- participants list of dictionaries where each dictionary in the list represents an individual student with the keys: participantID, userID, finalGrade, comments, annotations, handin.date, handin.ip, submission.paper.name, submissions.paper.size, submission.paper.downloadUrl and two empty columns. Two of the keys contained data buried deeper:
    - comments contained a list of dictionaries with data on comments where the keys were: id (assessorId), text (comment itself), timestamp, type, shareParticipant, shareAssessor, containerId, flowId, userId, fileId, asReviewer and explanationId.

- annotations contained a list of dictionaries with data on annotations where the keys were: id (assessorId), hash, pageNum, color, text, width, height, x, y, type, containerid, userid.

Revealing the contents of the json files we received required a process of unpacking detailed below. Having unpacked the data, questions regarding usability and GDPR were at the forefront. The flowData, assignment and assessor columns contained data unspecific to the rows with submission and grade that we primarily needed for our analysis. The assessor column also contained the name of assessors in plain text. We therefore decided not to include any data from these three columns in the datasets we were preparing for use in the analysis.

From the participants column we needed to include as a minimum the grade given and the actual submission text. There was also data available which we for GDPR reasons deliberately excluded from the datasets. Examples are participantId, student userId and in particular handin.ip.

The data we selected to proceed with was: Grade, Handindate, Size, timestamp of comment, userId of assessor extracted from the first comment dictionary item in the list of comments, comment text, submission title and submission text.

Data preparation

The code used for unpacking the data is detailed in the colab notebook appendix B. In summary we used the Spyder console and the python libraries pandas, html, re, requests, io, pypdf and sklearn to wrangle the jsons received from UNIwise into csv file's for uploading to DataRobot for further analysis. The wrangling consisted of the following main steps for each of the json files received:

- Flattening the nested structure of the json into a tabular format.
- Selecting the data to proceed with.
- Stripping html formatting from comments
- Dropping all rows without a link to submission text.

- Looping over all the rows with a function that opened the url linking to the submission pdf, extracted the text using Pypdf and saved the text to a new column in the dataframe.
- Saving the dataframe as a csv.

Json	assignment.name	subtitle	code
6639775	STR36053 Strategi - oppgave høst 2022.pdf	Term paper 100% - R	202220||20715||IN08||R||P
6639778	STR36053 Strategi - oppgave høst 2022.pdf	Term paper 100% - T	202220||20617||IN08||T||P
6639770	STR36053 Strategi - oppgave høst 2022.pdf	Term paper 100% - F	202220||20491||IN08||F||P
6639766	STR36053 Strategi - oppgave høst 2022.pdf	Term paper 100% - W	202220||20423||IN08||W||P
6639761	STR36053 Strategi - oppgave høst 2022.pdf	Term paper 100% - B	202220||20840||IN08||B||P
6450101	MRK3414Skriftlig eksamen h 2021.pdf	Take-home examination 30% - F	202120||22470||IN11||F||H
6450099	MRK3414Skriftlig eksamen h 2021.pdf	Take-home examination 30% - B	202120||22366||IN11||B||H
6450097	MRK3414Skriftlig eksamen h 2021.pdf	Take-home examination 30% - R	202120||22365||IN11||R||H
6450095	MRK3414Skriftlig eksamen h 2021.pdf	Take-home examination 30% - T	202120||22364||IN11||T||H
6450093	MRK3414Skriftlig eksamen h 2021.pdf	Take-home examination 30% - W	202120||22363||IN11||W||H
6445370	HIS 34106_202120_01.12.2021_14.00-17.00_QP.pdf	Written examination 100% - T	202120||22622||IN02||T||E
6445368	HIS 34106_202120_01.12.2021_14.00-17.00_QP.pdf	Written examination 100% - R	202120||22621||IN02||R||E
6445365	HIS 34106_202120_01.12.2021_14.00-17.00_QP.pdf	Written examination 100% - W	202120||22620||IN02||W||E
6445354	HIS 34106_202120_01.12.2021_09.00-12.00_QP.pdf	Written examination 100% - B	202120||20917||IN02||B||E
6445350	HIS 34106_202120_01.12.2021_09.00-12.00_QP.pdf	Written examination 100% - R	202120||20804||IN02||R||E
6445346	HIS 34106_202120_01.12.2021_09.00-12.00_QP.pdf	Written examination 100% - T	202120||20735||IN02||T||E
6445342	HIS 34106_202120_01.12.2021_09.00-12.00_QP.pdf	Written examination 100% - W	202120||20413||IN02||W||E
6386998	MRK 3414 Markedsføringsledelse Casetekst Oda Konte.pdf	Term paper 100% - W	202120||22068||IN11||W||P
6346624	MRK 3414 Markedsføringsledelse Casetekst Oda (002).pdf	Term paper 70% - R	202120||20809||IN11||R||P
6346603	MRK 3414 Markedsføringsledelse Casetekst Oda (002).pdf	Term paper 70% - T	202120||20741||IN11||T||P
6346582	MRK 3414 Markedsføringsledelse Casetekst Oda (002).pdf	Term paper 70% - F	202120||20643||IN11||F||P
6331234	MRK 3414 Markedsføringsledelse Casetekst Oda (002).pdf	Term paper 70% - W	202120||20441||IN11||W||P
6346657	MRK 3414 Markedsføringsledelse Casetekst Oda (002).pdf	Term paper 70% - B	202120||20923||IN11||B||P
6345085	STR36053 Strategi - oppgave høst 2021.pdf	Term paper 100% - B	202120||20935||IN08||B||P
6345079	STR36053 Strategi - oppgave høst 2021.pdf	Term paper 100% - R	202120||20821||IN08||R||P
6345067	STR36053 Strategi - oppgave høst 2021.pdf	Term paper 100% - T	202120||20753||IN08||T||P
6345058	STR36053 Strategi - oppgave høst 2021.pdf	Term paper 100% - F	202120||20650||IN08||F||P
6331885	STR36053 Strategi - oppgave høst 2021.pdf	Term paper 100% - W	202120||20478||IN08||W||P

As the list above shows we had multiple jsons with submissions to the same assignments, but from different campuses at BI. We concatenated all submissions belonging to the same assignments into six csv files, one for each course code. We did not include a column identifying campus although looking back it might have been interesting to explore how including this data would affect model accuracy.

The table below lists the size of the six csv files we prepared for uploading to DataRobot for further analysis.

STR36053H22	125 284 KB
STR36053H21	124 965 KB
MRK3414Oda	217 443 KB
MRK3414H21	26 832 KB
HIS34106H21F	14 934 KB
HIS34106H21E	12 861 KB

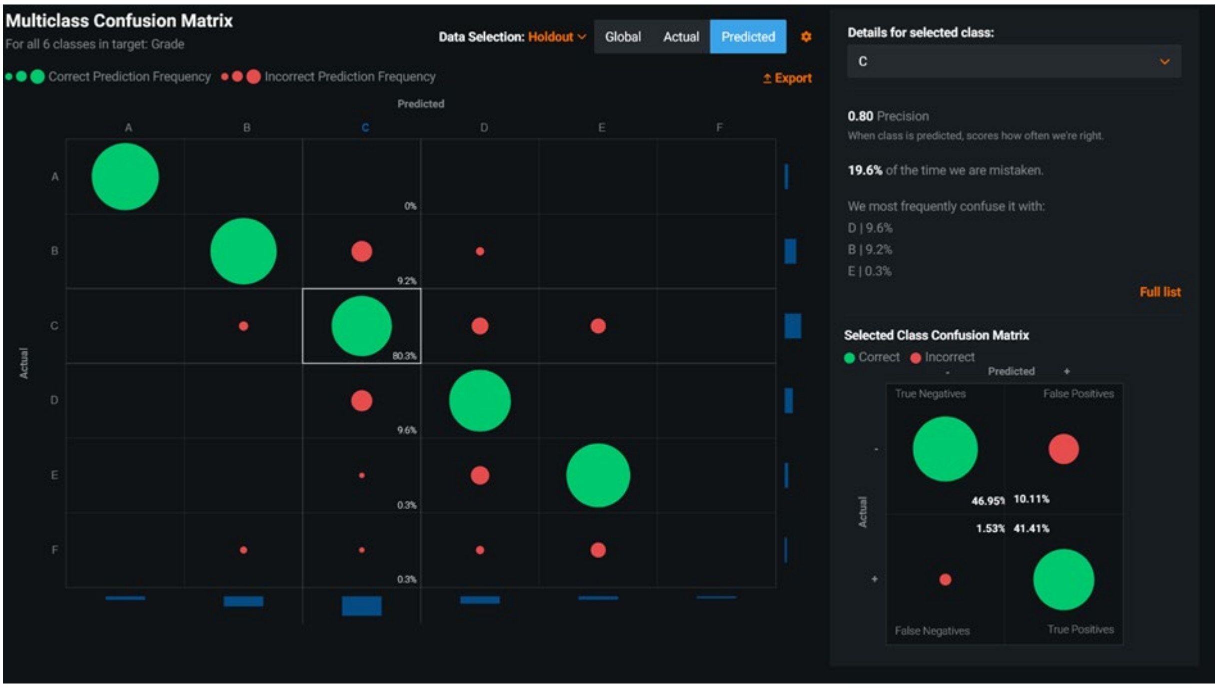
We considered doing preprocessing on the text data in Spyder such as stemming and stop word removal. Considering how operationalizing text data using TFIDF will discount the importance of a frequent word in the document if it is also frequent in the collection of documents, we were advised that stop word removal was not necessary. Stemming and other text preprocessing in Spyder might have improved results, but for the first iteration of analysis we decided to proceed without any work done to the text in Spyder, leaving DataRobot to work with the raw text.

Model training

Iteration 1 - Data leak discovery

The first dataset we uploaded to DataRobot was the 125MB csv file with data from the STR36053 exam given in the autumn of 2022. First, we created a subset of the features with only grade and text. We considered comments to be especially important to leave out assuming that adjectives in the comments might turn out to be a source of data leak. We selected grade as the target, let DataRobot keep all default settings including the recommended optimization metric LogLoss and started comprehensive training where DataRobot searches wide and deep for the best model to fit to the data.

The results were spectacular. DataRobot recommended a Keras Slim Residual Neural Network classifier that achieved a LogLoss of 0,37, 0,43 & 0,36 respectively on validation, cross validation and holdout. The confusion matrix below illustrates how good these results were.



The precision when predicting the grade, A was 100%, for C which was the clear majority class the precision was 80%. Only 0,6% of predictions were outside the grade bracket B-D showing that if the model misclassified a submission, at least it was not off by more than one grade either up or down.

We sent these results to Chandler Johnson who received them with polite suspicion commenting that they looked “uncanny good”. We set out looking for data leaks but were confounded since grade and text was the only data provided for training. After some investigations we realized that students were submitting in groups, but each individual was receiving a grade. In other words, there were multiple rows where the text and the grade were identical. Because the default partitioning setting is set to random the model was likely to see exactly the same rows in several cross-validation folds and in holdout. This was a major data leak that explained why we had achieved such outstanding performance. To solve this problem, we revisited Spyder to remove the duplicates from all six datasets before returning to DataRobot.

Iteration 2 – As good as it gets

The second dataset uploaded to DataRobot was the revised dataset from STR36053. By removing duplicates, the file size shrank from 125MB to 59MB. The performance after plugging the data leak deteriorated significantly. The precision in predicting grade C for instance worsened from 80% to 46,6%.

Believing we had solved the data leak problem we proceeded by fitting models to the five remaining datasets. For all of them we selected the feature subset grade and text, left default settings as is including the optimization metric LogLoss. The recommended model and its LogLoss on validation, cross validation and holdout for all six datasets are shown in the table below.

	LogLoss		
	Validation	Cross - V	Holdout
Random classification Benchmark	1,79	1,79	1,79
STR36053H21	1,249	1,316	1,337
STR36053H22	1,322	1,260	1,302
MRK3414Oda	1,141	1,189	1,220
MRK3414H21	1,208	1,191	1,189
HIS34106H21F	1,284	1,283	1,341
HIS34106H21E	1,291	1,270	1,218

We found that one model consistently outperformed the others and that model performance across the different datasets was fairly consistent hovering around LogLoss = 1,25.

Iteration 3 - Re-evaluating business use case

At this point it was clear that we would not achieve the model performance needed to support ideas for deployment in the direction of replacing human graders. Our focus then shifted towards one main idea for deployment; namely training on last year's exam to provide automated grade indications mid-term to next year's students after giving them last year's exam. The merits of this idea will be discussed in further detail later.

Proceeding with model evaluation given this use case we explored how accurately the model would indicate a student's grade. To be precise we set a bracket for each actual grade consisting of the grade above and below. For the actual grade A the bracket was set to A-B, for B it was A-C etc and for F it was E-F.

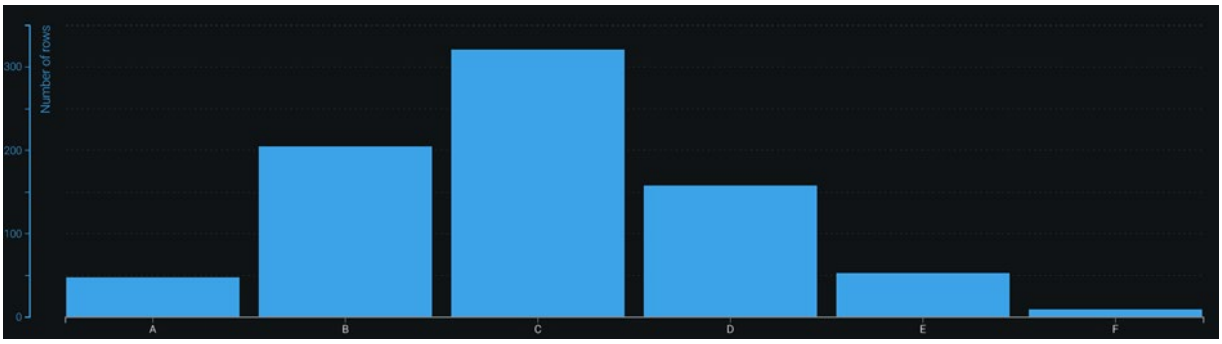
Since we had not created any partitions before loading data to DataRobot we had to revisit Spyder again and use train_test_split to save a "pre-holdout" partition to test running predictions. We obviously couldn't run predictions on the previously trained model since it would have trained on the data we wanted to test running predictions on. So, we retrained the model on 80% of the UniqueSTR36053H21 dataset. We chose this one for training run time, but it's worth noting that this was one of the datasets producing the worst performing models, now that we reduced the size of the training data performance dropped further to LogLoss of 1.26, 1.36, & 1.42 respectively for Validation, Cross Validation & Holdout. It's worth noting because we could expect to improve on the following results.

row_id	Grade	Bracket	Predictions in bracket	Prediction A	Prediction B	Prediction C	Prediction D	Prediction E	Prediction F
0	D	C-E	75 %	4 %	20 %	52 %	17 %	6 %	1 %
1	B	A-C	80 %	31 %	10 %	40 %	10 %	6 %	4 %
2	B	A-C	79 %	7 %	33 %	39 %	14 %	5 %	1 %
3	B	A-C	85 %	7 %	30 %	47 %	10 %	4 %	1 %
4	C	B-D	89 %	4 %	21 %	50 %	18 %	6 %	1 %
5	B	A-C	85 %	10 %	38 %	37 %	11 %	4 %	1 %
6	C	B-D	89 %	5 %	35 %	41 %	13 %	5 %	1 %
7	E	D-F	20 %	31 %	10 %	40 %	10 %	6 %	4 %
8	D	C-E	78 %	4 %	16 %	51 %	21 %	6 %	2 %
9	B	A-C	81 %	6 %	27 %	48 %	13 %	5 %	1 %
10	A	A-B	37 %	7 %	29 %	48 %	11 %	3 %	1 %
11	B	A-C	66 %	3 %	16 %	47 %	26 %	7 %	1 %
12	B	A-C	73 %	7 %	26 %	40 %	20 %	6 %	2 %
13	D	C-E	67 %	6 %	26 %	41 %	20 %	6 %	2 %
14	F	E-F	13 %	5 %	12 %	34 %	35 %	10 %	3 %
15	C	B-D	91 %	4 %	24 %	47 %	19 %	5 %	1 %
16	C	B-D	88 %	5 %	24 %	48 %	16 %	5 %	1 %
17	B	A-C	76 %	4 %	26 %	47 %	18 %	5 %	1 %
18	D	C-E	77 %	4 %	17 %	49 %	21 %	7 %	1 %

236	E	D-F	32 %	5 %	18 %	45 %	23 %	8 %	2 %
237	D	C-E	69 %	6 %	24 %	46 %	17 %	6 %	2 %
238	D	C-E	71 %	6 %	21 %	45 %	19 %	7 %	2 %
239	B	A-C	77 %	5 %	27 %	45 %	18 %	4 %	1 %
240	A	A-B	37 %	5 %	33 %	42 %	15 %	5 %	1 %
241	D	C-E	71 %	5 %	22 %	48 %	17 %	7 %	1 %
242	C	B-D	88 %	8 %	34 %	44 %	10 %	3 %	1 %
243	C	B-D	89 %	5 %	27 %	40 %	21 %	5 %	1 %
244	D	C-E	77 %	4 %	17 %	43 %	27 %	7 %	2 %
245	C	B-D	90 %	5 %	26 %	45 %	18 %	4 %	1 %
246	B	A-C	81 %	5 %	37 %	39 %	14 %	4 %	1 %

73 %	Total average in bracket
40 %	A average in bracket
78 %	B average in bracket
88 %	C average in bracket
72 %	D average in bracket
31 %	E average in bracket
12 %	F average in bracket

As we can see here and in all the confusion matrixes as shown in appendix A the models predict majority classes more accurately than minority classes. (With some exceptions possibly due to chance and few instances) The grade distribution for the model that made the predictions above look like this:



Down-sampling the majority classes or up-sampling the minority classes might help, but we did not explore these options. Down-sampling might be too costly given the limited number of rows. In case of up-sampling minority classes we'd have to be careful with how we partitioned our data not to cause any data leaks.

It's worth noting that the use of brackets above will inflate the model's perceived accuracy. This is because the majority classes favored by the classification models are present in more brackets since they are centered in the distribution of grades.

Looking at the results we believe the accuracy as presented above is insufficient for deployment, but that the idea merits further work towards the deployment of an automated grading guidance tool. Ie auto grading of last year's exams given to next year's students as a guidance tool for indicating to each student how well they are performing in each class during the semester.

Iteration 4 – Importance of data quantity

The wording in the STR36053 assignment given in autumn of 2021 and 2022 respectively was identical. This was overlooked initially, but discovering the fact prompted us to test how combining the two datasets and retraining would affect model performance.

There were 979 submissions to the assignment in autumn of 2021 and 1008 submissions in autumn of 2022. The submissions from both years were assessed by the same three assessors as indicated by the unique userId data column.

The table below shows that the improvement on validation and cross-validation after combining the datasets was insignificant. Doubling the training data seems to have made the model somewhat more robust as we can see holdout performance improving slightly. This means it generalized better than the individually trained models.

	LogLoss		
	Validation	Cross - V	Holdout
Autumn 21	1,249	1,316	1,337
Autumn 22	1,322	1,260	1,302
Combined	1,277	1,278	1,253
Average of H21 & H22	1,286	1,288	1,319
Improvement	-0,01	-0,01	-0,07

The main business use case proposition given the model accuracy we've managed to achieve is to provide automated grading as a service to students during the course as a guidance tool giving them an indication of their performance. If BI was willing to invest in grading these exams for a few years in addition to the actual exams, they might acquire a dataset with submissions large enough for the given assignment to eventually train a fairly accurate model. This would however run counter to the business objective of cutting costs associated with grading. Rather the solution has to be to look for existing data. We have made an effort to find the largest possible datasets for use in our analysis. Datasets with the most submissions to equally worded assignments. Hower we cannot say for sure that we've found the best data to work with and would recommend searching the organization for even bigger datasets to work with should someone pursue this work further.

The analysis provided, with the current data provided, will in any means be minor, compared to the amount of data that models such as ChatGPT have trained on. It could be argued that data analysts could with more data and resources provide a significantly refined and trained model.

Iteration 5 – Audit use case

The audit use case idea was as follows. If we could improve model accuracy by providing data such as timestamp of comments, assessorId and campus code then that would indicate unequal grading of exams. Given a deployment where the grading of many assignments was audited this way you could compare for which assignments providing this extra data improved model accuracy the most or simply set a threshold for model improvement over which a flag would be raised, and human resources would be allocated to audit the grading of the assignment. Given that resources are already allocated for this type of audit it could improve the allocation of these resources. We think the method would be particularly exposed to risks of overfitting.

The analysis did not yield any findings. We retrained the largest of our datasets; UniqueMRK3414Oda with timestamp of assessor comment and assessorId in addition to text and grade. Kept all settings equal and ran a comprehensive search for the best model fit. Logloss changed from 1.14, 1.18 & 1.21 to 1.14, 1.17 & 1.24 for validation, cross-validation and holdout respectively. Performance on the validation folds remained stable while performance on holdout deteriorated slightly. Likely due to some overfitting on the new data that did not generalize.

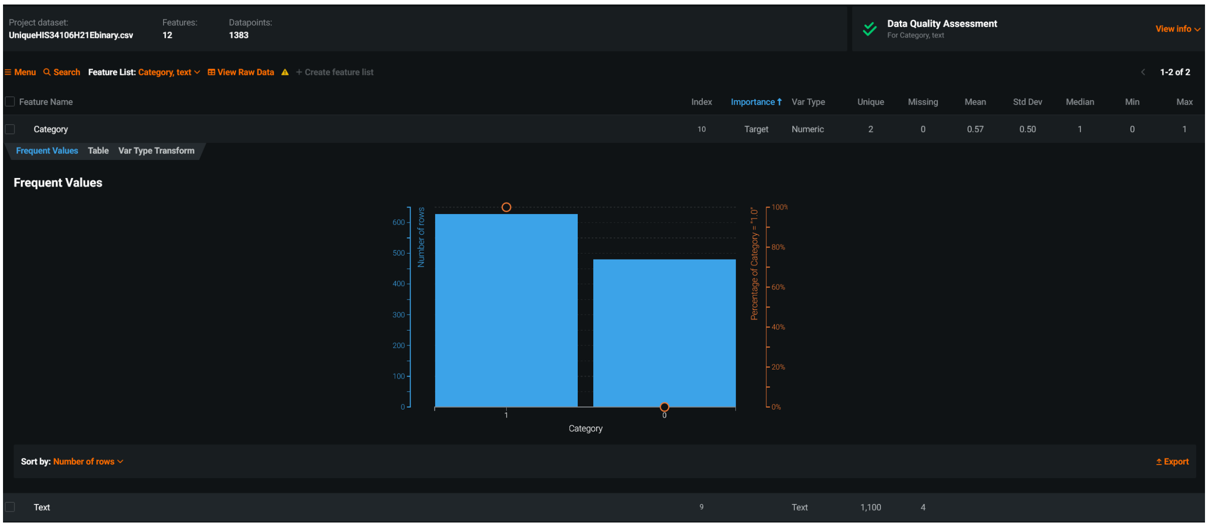
It's worth noting that the timestamp and assessorId in our dataset included only the first list item of dictionary entries with comments. This was a shortcut we did because we didn't expect to prioritize using this data. In the "Oda" dataset for instance there were only 539 timestamps out of a total of 1375 unique exam submission texts.

In conclusion we believe the audit use case might be worth pursuing further, but our analysis has been flawed by poor data preparation for this use case and we decided to not prioritize exploring this further.

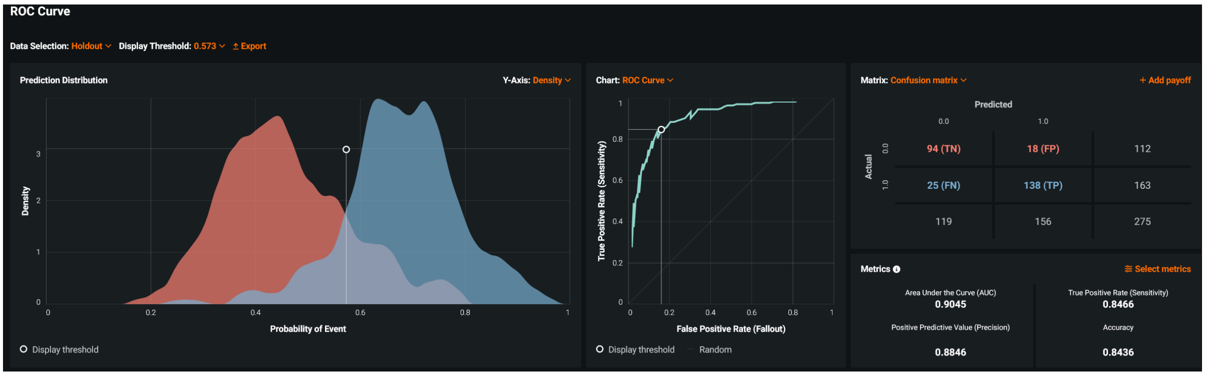
Iteration 6 – Binary classification

Chronologically we came back to this iteration after having written the "Evaluation metrics" chapter below.

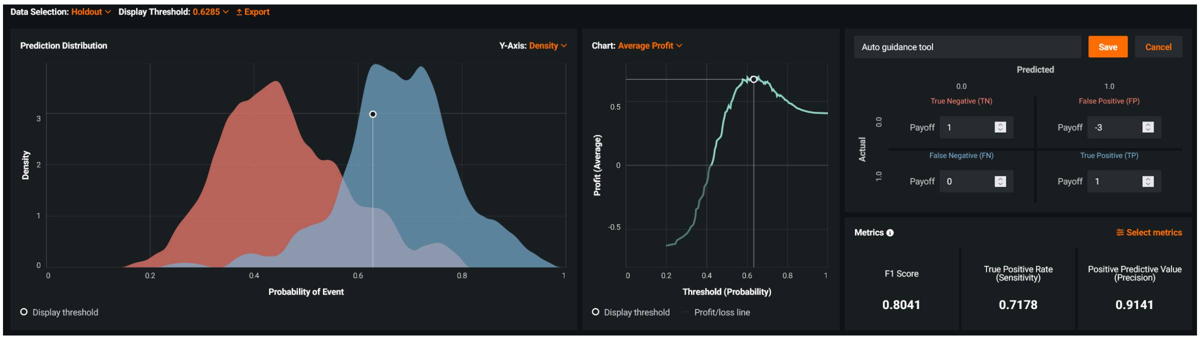
Noting that HIS34106H21E seemed to be the dataset used to train the best performing model, likely because it was the least unbalanced dataset. We wanted to iterate back to data preparation again and see how well this dataset would perform on a binary classification problem predicting if the grade was A,B,C or D,E,F. Having prepared the data, we got fairly balanced classes with 1 representing the best grades and 0 the worst grades:



Not surprisingly the binary classification model performed much better than the multiclass class probability estimation models. The best performing model was an Elastic-Net Classifier (L2 / Binomial Deviance) achieving LogLoss of 0,497 & AUC of 0,905 on holdout. As we can see in the prediction distribution below the two classes are quite separable.



In the profit matrix shown below we set the "profit" of indicating correctly to a student whether or not they are performing at a C level or better to 1(TP or TN). We set the profit of falsely predicting that a student is performing at a D level or worse to 0 (FN) assuming that the possible outcomes of motivating or demotivating the student to study are equal. Finally, we set the profit of falsely predicting that a student is performing at a C level or better when in fact they are performing at a D level or worse to –3 (FP) penalizing this heavily assuming the outcome would be for the student to relax their efforts. We then get the profit graph below showing that deploying the model will have a positive aggregate outcome for students given the confusion and profit matrix values.



As it is not included in appendix A. Here is the word cloud for this classification model:

Model description

Looking into the multiclass classification model which was recommended for all of our datasets we asked ChatGPT-4 the following:

"We have used DataRobot to train on seven different datasets with written exams and grades. In our supervised learning with text based multiclass probability estimation project DataRobot found the same model for all seven datasets. The model it choose was Keras Slim Residual Neural Network Classifier using Training Schedule (1 layer: 64 units) M66 BP18 . Why do you think this model outperforms all the others every time even though 25 models were tested for each of the seven datasets?"

This was part of the answer from ChatGPT-4:

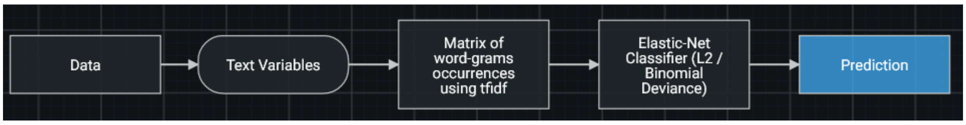
"...

1. *Residual connections: Residual Neural Networks (ResNets) are known for their ability to mitigate the vanishing gradient problem by introducing skip connections between layers. This allows gradients to flow more easily through the network, making it possible to train deeper models and achieve better performance.*
2. *Appropriate model complexity: The Keras Slim ResNet with 1 layer and 64 units might strike a good balance between model complexity and generalization for your datasets. A model that is too complex can overfit the data, while a model that is too simple might not be able to capture the underlying patterns in the data.*

..."

We clearly could not have given a better answer. We're allowing ourselves some leeway with using Chat GPT as we find it fitting and somewhat amusing to include ML generated text evaluating why an ML model designed to capture patterns in text data is the chosen model for our ML project.

Model info, blueprints and neural network visualizer for these ResNets are shown in appendix A. For the binary classification problem addressed in iteration 6 an Elastic-Net Classifier won out. Below is the blueprint for that model.



Model evaluation

Evaluation by ChatGPT

Tempted by the helpful ChatGPT to keep engaging it, we posed the following question and got the subsequent answer:

Us: “We are averaging around 1.25 logloss on holdout for the models we've trained. How would you evaluate the model performance?”

Chat GPT-4: “A log loss (also known as logarithmic loss or cross-entropy loss) of 1.25 on the holdout set provides some information about the model's performance, but it's important to put it in context. Log loss is a measure of how well the predicted probabilities match the true labels, with a lower value indicating better performance.

To evaluate the model performance, consider the following:

1. *Compare with a baseline: To determine if the log loss of 1.25 is good or not, you should compare it to a baseline model. A common baseline for multiclass problems is the "random guess" model, which assigns an equal probability to each class. For a problem with K classes, the random guess model would have a log loss of -log(1/K). If your model has a significantly lower log loss than this baseline, it's a sign that it's learning meaningful patterns in the data.*
2. *Relative performance: Compare the performance of your model with other models that were trained on the same dataset. If your model has the lowest log loss among all the models, it indicates that it's performing relatively well. However, if there are other models with similar log loss values, it might be worth considering them as well, as they may offer other benefits, such as faster training or inference times.*
3. *Performance on other metrics: Log loss is just one metric to evaluate a model's performance. For multiclass classification problems, you might also want to consider other metrics, such as accuracy, F1-score, or confusion matrix, to gain a more comprehensive understanding of how well your model is performing.*
4. *Performance in the context of the problem: The acceptable level of log loss depends on the specific problem you're trying to solve. In some cases, a log loss of 1.25 might be acceptable, while in others, it might be too high. Think about the implications of the*

model's predictions in the context of the application and decide whether the current performance is sufficient.

5. *Evaluate the model on new, unseen data: To get a better sense of the model's true performance, evaluate it on new data that it has not seen during training or validation. This can help you gauge how well the model generalizes to new examples and whether it's prone to overfitting.*

In summary, to evaluate your model's performance, compare its log loss to a baseline, consider other performance metrics, and evaluate it in the context of the specific problem and application."

Us: "Our model has six classes"

ChatGPT-4: "If your problem has 6 classes, then the random guess model would assign an equal probability to each class, which is 1/6 or approximately 0.167 for each class. The log loss for the random guess model can be calculated as -log(1/6), which is approximately 1.79.

Since your model has a log loss of 1.25 on the holdout set, it performs better than the random guess model, as the log loss of your model is lower than that of the random guess model (1.25 < 1.79)."

Following the checklist provided by ChatGPT-4; the Keras Slim Residual Neural Network is both outperforming the baseline *monkey with a dart* method and all the other models trained. Whether it will serve to be useful is the question.

Evaluation metrics

DataRobot provided the evaluation metrics AUC, Accuracy, Balanced Accuracy, FVE Multinominal and LogLoss for our trained models. In the results summarized below conditional formatting indicates which dataset performed best for each metric on each data selection:

	LogLoss			AUC			Balanced Accuracy			FVE Multinominal		
	Validation	Cross - V	Holdout	Validation	Cross - V	Holdout	Validation	Cross - V	Holdout	Validation	Cross - V	Holdout
Random classification Benchmark	1,79	1,79	1,79	0,5	0,5	0,5	0,167	0,167	0,167	0	0	0
STR36053H21	1,249	1,316	1,337	0,668	0,653	0,666	0,214	0,214	0,248	0,089	0,084	0,104
STR36053H22	1,322	1,260	1,302	0,621	0,661	0,603	0,210	0,232	0,201	0,050	0,089	0,047
STR36053all	1,277	1,278	1,253	0,673	0,661	0,707	0,218	0,242	0,220	0,108	0,098	0,119
MRK3414Oda	1,141	1,189	1,220	0,718	0,709	0,717	0,332	0,309	0,307	0,175	0,169	0,155
MRK3414H21	1,208	1,191	1,189	0,707	0,714	0,735	0,259	0,283	0,302	0,139	0,152	0,166
HIS34106H21F	1,284	1,283	1,341	0,795	0,770	0,753	0,377	0,376	0,347	0,235	0,228	0,182
HIS34106H21E	1,291	1,270	1,218	0,783	0,781	0,777	0,374	0,371	0,390	0,230	0,238	0,256

We've left accuracy out as it does not perform well with imbalanced datasets. (DataRobot, 2023).

The benchmark LogLoss was explained by ChatGPT above. The explanation for AUC, Balanced Accuracy and FVE Multinominal random baseline values by ChatGPT-4 are as follows:

"For a multiclass classification problem with six classes, the balanced accuracy can also be calculated by considering an equal probability for each class. Balanced accuracy is the average of recall (sensitivity) obtained on each class. In the case of random guessing, the recall for each class would be the same, as we would be assigning equal probability to each class.

Since there are six classes, the probability of randomly guessing the correct class is 1/6 or approximately 0.167 (16.7%).

In this case, the balanced accuracy would also be the same as the random guess baseline accuracy, which is approximately 16.7%. This is because the recall for each class would be the same, and we're taking an average of equal values."

"For a random classifier, the Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) curve would be 0.5 for each class in a one-vs-all (or one-vs-rest) scenario. This is because a random classifier has no discrimination ability, and its true positive rate (sensitivity) would equal its false positive rate at every decision threshold, resulting in a diagonal ROC curve.

Since the AUC value is the same for each class in a random classification (0.5), the weighted average AUC value would also be 0.5, regardless of the class frequencies. This is because the weights will sum up to 1, and the AUC value for each class is the same.

So, for a random classifier in a multiclass classification problem with six classes, the weighted average AUC value would be 0.5.

"Fraction of Variance Explained (FVE) is a metric used to evaluate the goodness of fit of a model in comparison to a baseline model. In the case of the multinomial distribution, FVE multinomial measures the deviance reduction achieved by the model in relation to the deviance of the baseline model.

For a random classifier in a multiclass classification problem, the predictions would be random and not expected to provide any reduction in deviance compared to the baseline model. The FVE multinomial can be calculated using the following formula:

FVE_multinomial = 1 - (Deviance of the model / Deviance of the baseline model)
Since we expect no reduction in deviance for a random classifier, the deviance of the model should be approximately equal to the deviance of the baseline model. Therefore:
FVE_multinomial = 1 - (Deviance of the baseline model / Deviance of the baseline model)
FVE_multinomial = 1 - 1
FVE_multinomial = 0

So, for a random classifier in a multiclass classification problem with six classes, the FVE multinomial value would be 0, indicating that the random classifier provides no improvement over the baseline model."

We suspect that there is potential to improve on these results as we will discuss further in the model tuning section below. We will therefore not be spending time evaluating how useful grade predictions or grade indications would be for students given this model accuracy.

Confusion matrix inspections

Looking at the confusion matrixes in appendix A the most striking finding is that misclassifications are tending strongly towards adjacent grades. This seems to indicate that the models are on to something.

Although obvious it is also noteworthy that the precision and recall for the minority classes are significantly worse than for the majority classes.

Word cloud inspections

There are word clouds for each grade for all the recommended multiclass classification models in appendix A. They read A through F top to bottom and left to right. The color gradient is a visualization of the word cloud coefficient value. We couldn't explain how it works if we tried, but basically it is a measure of how a wordgram or n-gram in the word cloud correlates to the prediction of the given grade. The stronger the color the stronger the correlation. Red indicates a positive correlation to the given grade and blue indicates a negative correlation to the given grade.

It is interesting to note how higher grades tend to be predicted more based on positive correlation to words or n-grams, while lower grades tend to be predicted more based on negative correlations as with this example from the UniqueMRK3414Oda dataset:

Grade A

Grade F

Stemming seems not to have been performed by DataRobot on the text data. In the word cloud for UniqueSTR36053H21 grade A we find "Jobb" "Jobbet" & "Jobben" as separate tokens with negative correlations to the grade.

Stop word removal might be unnecessary given that the TFIDF method discounts the importance of frequent words or n-grams in a document if the word or n-gram is also frequent in the collection of documents. This seems to be true as we could not find any typical stop words in the word clouds in appendix A.

We find numbers with high predictive value in most if not all the datasets. In some, such as the word cloud for UniqueMRK3414Oda grade D the model finds the number 37 to have the strongest negative correlation to that particular grade of all tokens in the dataset. This could be a case of overfitting.

The word clouds are dominated by single words and numbers with a few two-word combinations interspersed.

We looked closer at the "Veiledning for sensur av STR36053"; the assessor grading guidelines and the wordcloud for this dataset to see if the model had picked up on some core terms sought after by the grading guidelines. The comparison was superficial, but we did make some findings. "Bransje" for instance was highlighted as the strongest positive predictor in the word cloud for grade A. This word is also mentioned eight times in the grading guidelines. It's interesting that the conjugation of strategy; strategic was highlighted in the word cloud for grade and mentioned 12 times in the grading guidelines. While the word strategy was mentioned 38 times in the grading guidelines but was completely absent from all word clouds. Perhaps an argument against stemming.

Model tuning

We have not made any attempts at tuning our models. Given that we are amateurs in the field we find it fairly safe to assume that model performance with the same datasets could be improved on by a professional data analyst. Our findings above might suggest that DataRobot did a limited amount of text pre-processing. Examples are findings of stemming not being done and numbers being included in the dataset. Moving forward, it would be interesting to test how different text pre-processing and hyperparameter tuning would affect model performance. Seeing how model accuracy would be affected by setting n-grams to 3 or more would also be very interesting.

Model deployment

None of the models trained are ready for deployment. We would recommend continuing with iterations of data preparation, model training and evaluation preferably by an expert data analyst until the improvements in model performance between iterations flatten out.

That being said, of the models we have trained we would select the binary classifier trained in iteration 6 for deployment. Deploying such an unpolished model would not warrant much data engineering. It might be interesting to gauge how an auto grading guidance tool would be received by the students before investing more time and money into model improvements. If so, we would recommend a strictly manual deployment of the model. A resource with the required skills could use the code provided in appendix B to wrangle the data in the jsons from UNIwise to prepare data for model training in the desired course. Then either licenses for DataRobot would have to be acquired or if the resource had skills to train a model in python, then the costs could be reduced. The platform for administering the exams could be WiseFlow. Or to avoid some wrangling of the dataset on which to make predictions, one could use a form builder tool to capture the text in tabular format directly.

How the results should be presented to the students might be affected by a more thorough analysis of appropriate values for the profit matrix and consequently where the threshold between positive and negative classification should be set. If the threshold was set to 0,5 then all predictions over 0,5 would be presented as “student is likely in the bracket A-C" and under would be presented as “student is likely in the bracket “D-F” We think the most informative and transparent way to present the results to students would be to explain that a number closer to 1 indicates a good performance while a number closer to 0 indicates a poor performance. Or simply that the number represents the probability of having a grade in the A-C bracket, the positive class. Here is an example of predictions made by the Elastic-Net Classifier (L2 / Binomial Deviance) on a partition of the UniqueHIS34106H21Ebinary dataset. (Note that the partition was made after the model was trained, so there is data leak in these predictions)

Row	Pred.	Label	Row	Pred.	Label	Row	Pred.	Label	Row	Pred.	Label	Row	Pred.	Label
0	0,32	0	11	0,27	0	22	0,77	1	33	0,41	0	44	0,33	0
1	0,76	1	12	0,78	1	23	0,79	1	34	0,26	0	45	0,72	1
2	0,53	1	13	0,38	0	24	0,32	0	35	0,83	1	46	0,70	1
3	0,54	1	14	0,51	1	25	0,39	0	36	0,30	0	47	0,38	0
4	0,67	1	15	0,76	1	26	0,85	1	37	0,86	1	48	0,69	1
5	0,63	1	16	0,76	1	27	0,39	0	38	0,78	1	49	0,32	0
6	0,51	1	17	0,23	0	28	0,30	0	39	0,32	0	50	0,65	1
7	0,36	0	18	0,52	1	29	0,81	1	40	0,51	1			
8	0,21	0	19	0,31	0	30	0,32	0	41	0,82	1			
9	0,32	0	20	0,58	1	31	0,24	0	42	0,40	0			
10	0,47	0	21	0,86	1	32	0,73	1	43	0,80	1			

To increase the value provided to students by the grading guidance tool one could present them with prediction explanations together with the prediction itself. Computing prediction explanations on these text-based classification problems have in our experience been time consuming work for DataRobot. If used commercially by BI, computing the explanations might not be worth it. If so; one could consider simply providing the word cloud.

If either after a manual deployment resulting in positive feedback from students or without such an experiment BI decides to invest in improving the predictive models we've started work on and the work produces models that satisfy specified requirements for accuracy, then a proper deployment will be needed.

Perhaps the most seamless alternative for deployment would involve partnering with UNIwise to provide feedback directly through their platform. It might be beneficial from a data engineering perspective. We are not really qualified to make recommendations concerning proper model deployment, so we returned to ChatGPT:

"Assuming we have a well-tested model ready for deployment, what are the hallmarks of a successful deployment?

A successful deployment of a machine learning model is characterized by several hallmarks. Ensuring that your deployment meets these criteria will help you achieve a smooth integration of your model into production and maximize its value to your organization:

1. *Seamless integration: The deployed model should be easy to integrate with your existing systems, applications, or services. This can be achieved using APIs, RPC mechanisms, or other communication protocols that ensure compatibility and minimize disruptions to existing workflows.*
2. *Low latency and high throughput: The deployed model should respond quickly to requests and be able to handle a large volume of requests without significant degradation in performance. This is especially important for real-time or near-real-time applications where low latency is critical.*
3. *Scalability: The deployment infrastructure should be capable of handling fluctuations in demand and scale up or down as needed. This can be achieved using cloud services, containerization, or other architectural patterns that enable dynamic resource allocation.*
4. *Robustness and reliability: The deployed model should be resilient to failures, errors, or unexpected inputs. This can be achieved through thorough testing, validation, and monitoring, as well as implementing fallback mechanisms or redundancy to ensure continued operation in the face of failures.*
5. *Security and privacy: The deployment should ensure that your model and data are protected from unauthorized access and comply with relevant data protection regulations. This can be achieved through access controls, encryption, and data anonymization techniques.*
6. *Monitoring and maintenance: A successful deployment should include continuous monitoring of the model's performance, as well as regular updates and retraining to maintain its accuracy and relevance. This may involve setting up automated monitoring systems, alerting mechanisms, and processes for updating the model with new data or improved algorithms.*
7. *Adaptability and flexibility: The deployment should be adaptable and flexible, allowing for easy updates and changes to the model or its parameters. This can be achieved through modular design, version control, and clear documentation.*
8. *Cost-effectiveness: A successful deployment should balance performance and resource requirements with cost considerations. This may involve optimizing the model or infrastructure to minimize resource usage or selecting cost-effective deployment options.*

9. *User satisfaction: Ultimately, a successful deployment should meet the needs of its users and provide value to the organization. This can be measured through user feedback, adoption rates, and the impact of the model on key performance indicators (KPIs) or business objectives.*

By ensuring that your deployment meets these hallmarks, you can maximize the value of your machine learning model and achieve a successful integration into your organization's systems and processes."

Conclusions and further recommendations

A great idea does not always provide the desired result. It is all part of the learning process; testing, adjusting and trying again. Machine learning can be a powerful tool for predicting grades on text-based term papers. However, there are several factors that can cause some models to fail while others work.

Business schools should start exploring the use of machine learning for grading term papers and other exams as we've argued under internal analysis, industry research and academic research. One of the main benefits is that it could help reduce the workload of faculty members and provide more consistent and objective grading. While automation of operations is increasing, machine learning automation projects for predicting grades still needs to go hand in hand with human assessment as a combination to ensure accurate and fair evaluations while providing personalized feedback to students. Automated grading systems should therefore be used as a tool to support the evaluation process, and faculty members should still review and provide feedback on student work.

To proceed with implementing machine learning for grading, BI should explore opening up for extracting large and diverse datasets of term papers, other exams and the grades. This is an investment and for that reason we should go through a cost-benefit analysis. Findings recommend continuing with iterations of data preparation, model training and evaluation preferably by an expert data analyst until the improvements in model performance between iterations flatten out. If, or perhaps rather when, ML grading of exams becomes a real option, then looking at how the exam questions themselves are framed and formulated will be important. A key to success is that faculty members should all be trained in how to interpret and use the automated grading results to provide feedback and support for their students.

Additionally, BI should establish guidelines for how the automated grading system should be used and ensure that it is fair and transparent for all students.

Why should BI continue to investigate the opportunities that lie in Machine learning for predicting grades? Because the opportunities are nearly endless. The grading- indication tool simplified to a binary classifier, with the necessary refinements, can be a great steppingstone considering that a manual deployment would be harmless, cheap and useful in addition to allowing for incremental improvements. It could be the first step towards a big lead in using ML for grading text-based exams.

Bibliography

Abdous, M'Hammed (2023). *How AI Is Shaping the Future of Higher Ed*, Inside Higer Ed https://www.insidehighered.com/views/2023/03/22/how-ai-shaping-future-higher-ed-opinion

Baker, R. S., & Yacef, K. (2009). *The state of educational data mining in 2009: A review and future visions.* Journal of Educational Data Mining, 1(1), 3–17.

Bujang, S.D.A. et al., (2021). *Multiclass Prediction Model for Student Grade Prediction Using Machine Learning*, in IEEE Access, vol. 9, pp. 95608-95621, 2021, doi: 10.1109/ACCESS.2021.3093563

Cerratto Pargman, T., Lindberg, Y. & Buch, A. (2023). *Automation Is Coming! Exploring Future(s)-Oriented Methods in Education*. Postdigit Sci Educ 5, 171–194.

DataRobot (2023), *DataRobot metrics, Accuracy/Balanced Accuracy.* Accessed 03.05.2023. https://app2.datarobot.com/docs/modeling/reference/model-detail/opt-metric.html#accuracybalanced-accuracy

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). *Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil.* Journal of Business Research, 94(February 2018), 335–343.

Gallagher, M., Breines, M. (2023). U*npacking the Hidden Curricula in Educational Automation: A Methodology for Ethical Praxis*. Postdigit Sci Educ 5, 56–76.

Gibson, P. (2023). *Enacting Empowerment Through an Automated Teaching Event: A Posthuman and Political Perspective*. Postdigit Sci Educ 5, 77–99.

Iatrellis, O., Savvas, I.K., Kameas, A. et al. (2020). I*ntegrated learning pathways in higher education: A framework enhanced with machine learning and semantics.* Educ Inf Technol 25, 3109–3129.

Mahboob, K.; Adif, R.; Haider, N. G. (2023). Q*uality enhancement at higher education institutions by early identifying students at risk using data mining. Mehran University* Research Journal of Engineering and Technology, [S.l.], v. 42, n. 1, p. 120-136, jan. 2023. ISSN 2413-7219.

Provost, F., & Fawcett, T. (2013). *Data science for business: What You Need to Know about Data Mining and Data-Analytic Thinking* (1st edition.). O'Reilly.

Sadler, M.; Regan, N. (2019). *Game changer. AlphaZero's Groundbreaking Chess Strategies and the Promise of AI. Alkmaar. The Netherlands.* New in Chess.

Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A. et al. (2023). *Making Sense of the Digital Automation of Education.* Postdigit Sci Educ 5, 1–14.

Strogatz, S. (2018). *One giant step for a chess-playing machine.* New York Times, 26.

Yağcı, M. (2022). *Educational data mining: prediction of students' academic performance using machine learning algorithms.* Smart Learn. Environ. 9, 11.

Appendix A - Data and model info…

Appendix B - Notebook

Appendix C – Exam texts

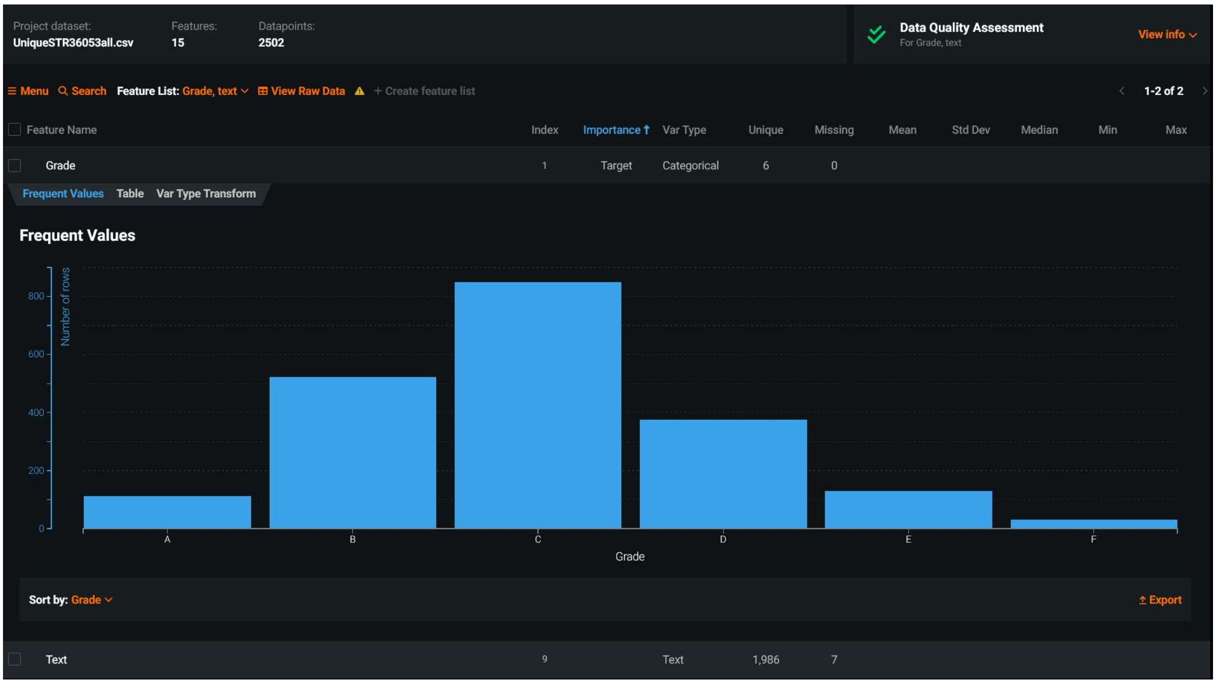
Content Appendix A

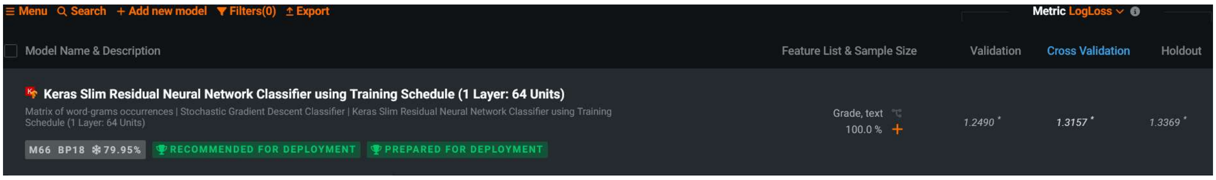
Appendix A

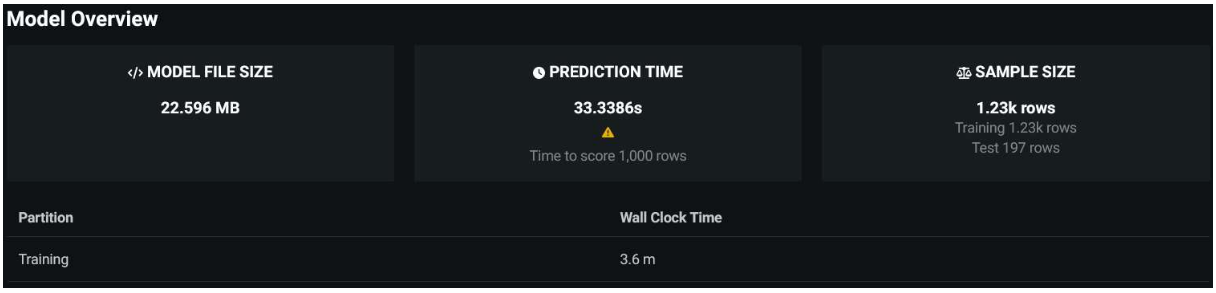
Data and model info, confusion matrixes & word clouds for each of the seven datasets trained on.

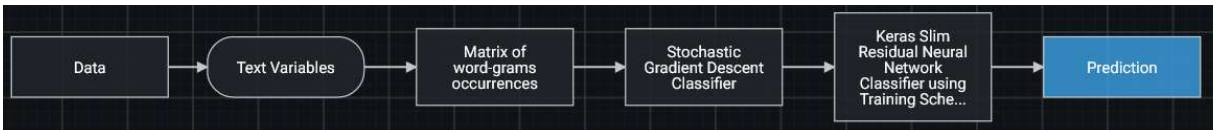
(Word clouds are for classes A-F read from top down & left to right. Words colored in red means positive correlation and blue color means negative correlation to the class in question.)

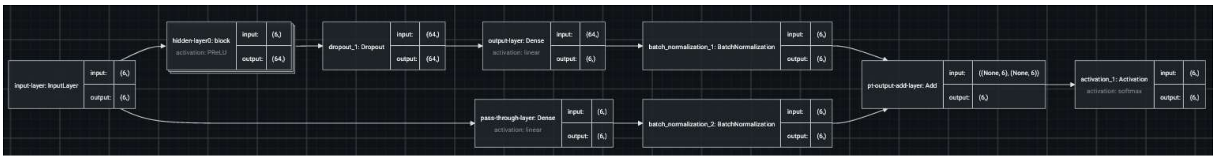
UniqueSTR36053H21.csv data and model info



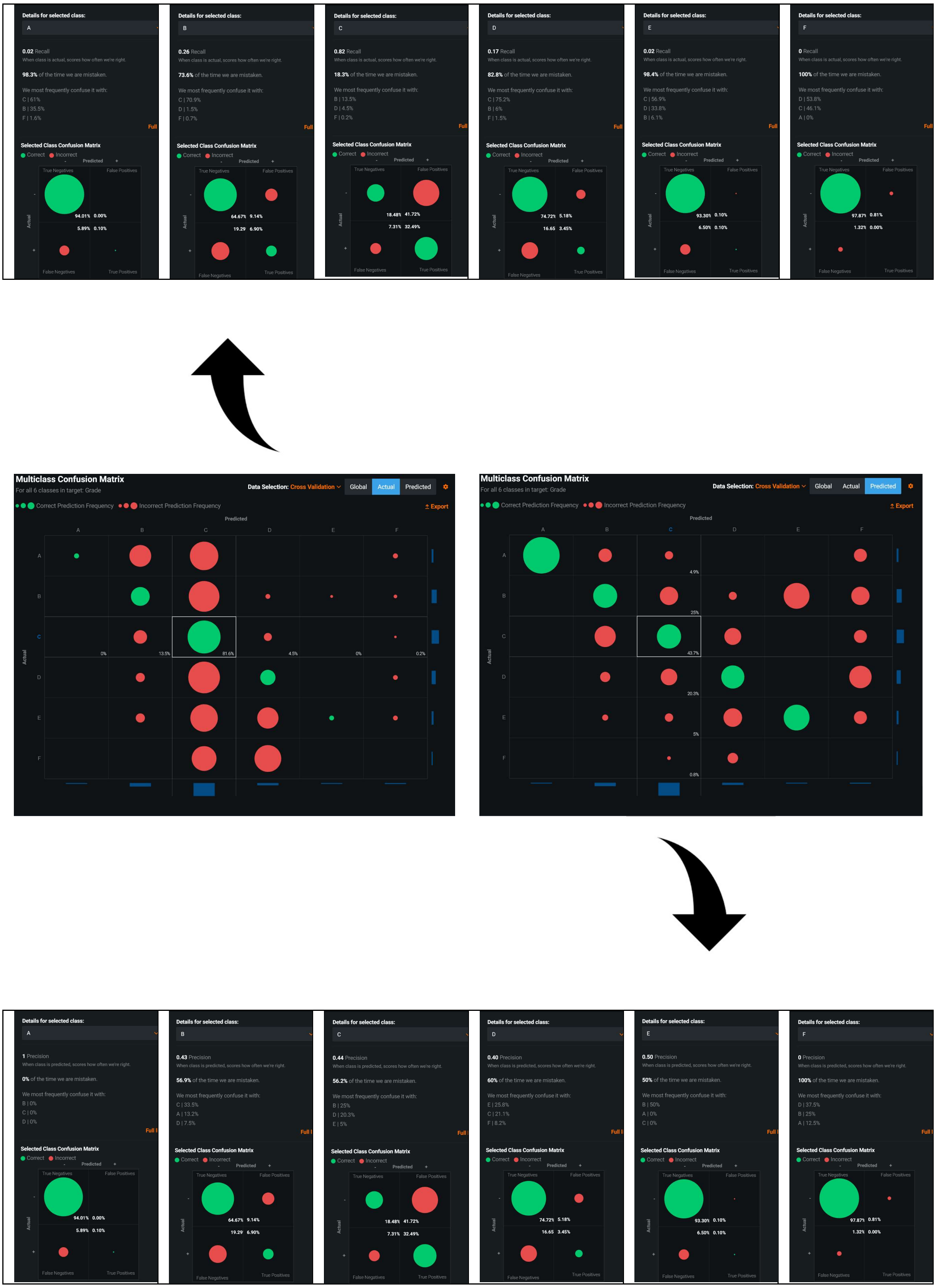






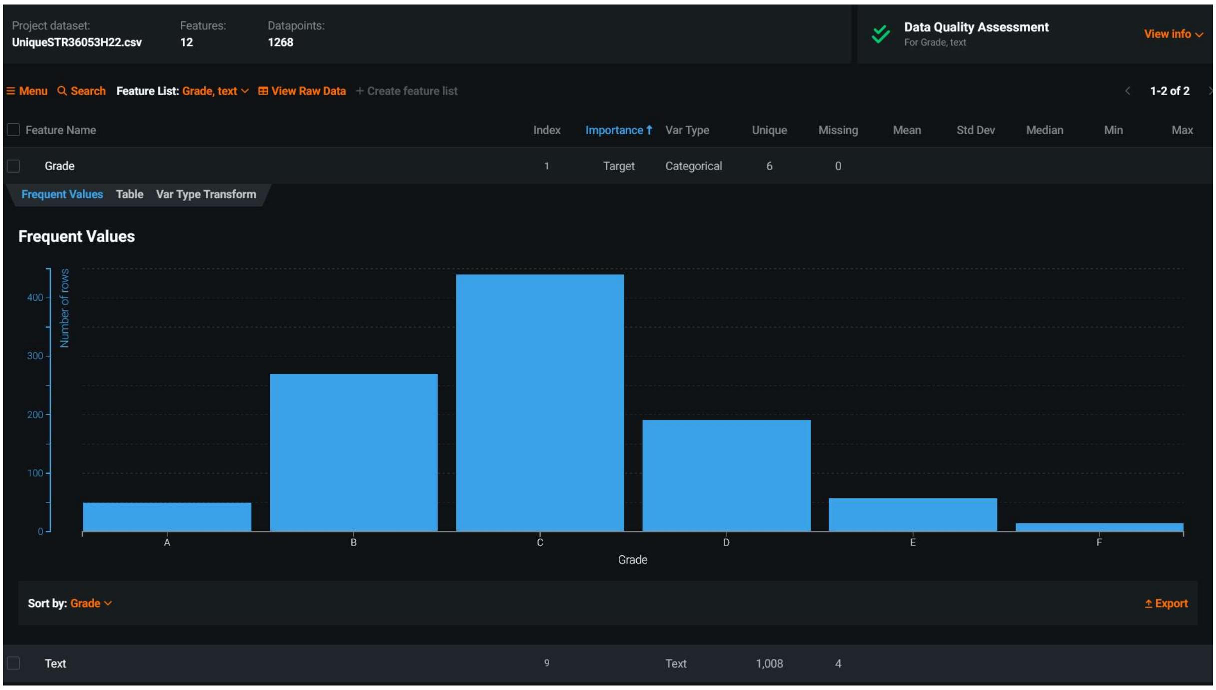


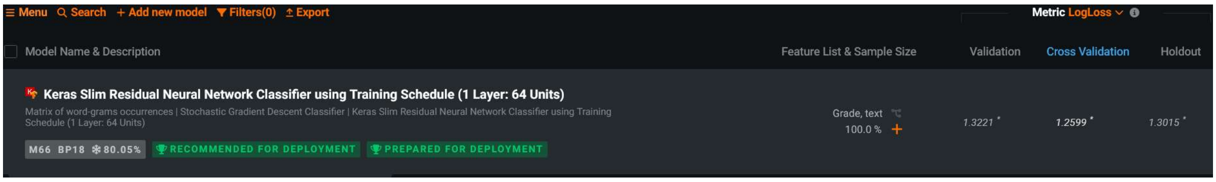
UniqueSTR36053H21.csv confusion matrix

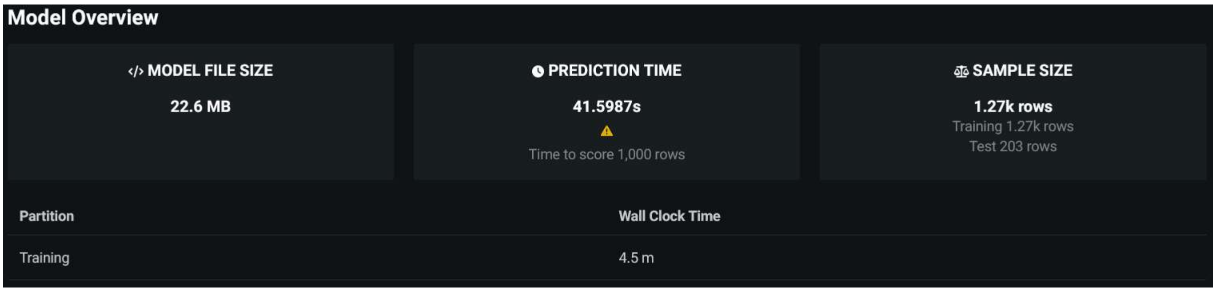


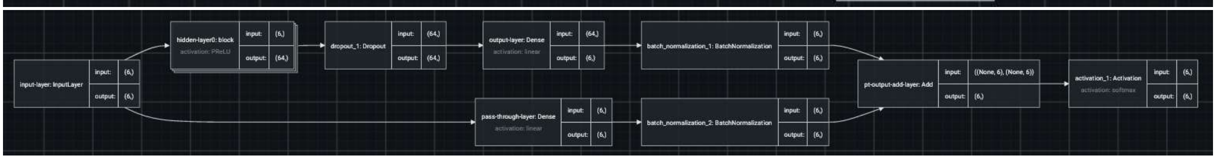
UniqueSTR36053H21.csv word clouds

UniqueSTR36053H22.csv data and model info

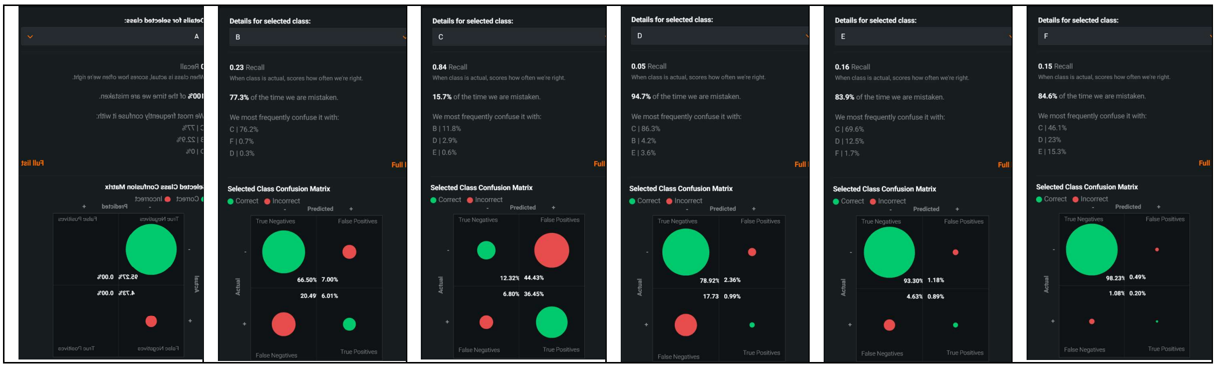


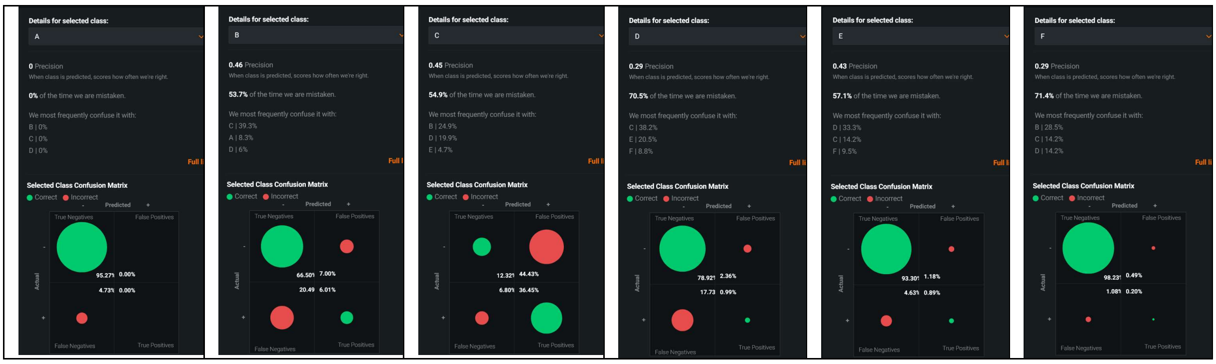






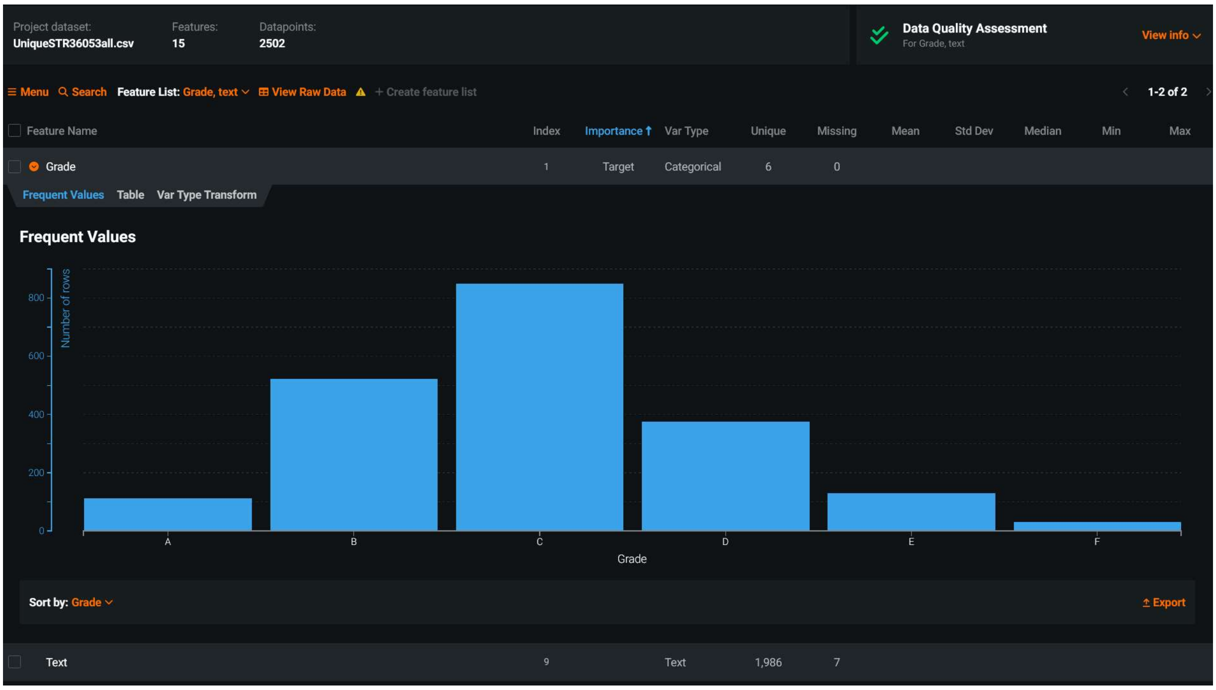
UniqueSTR36053H22.csv confusion matrix

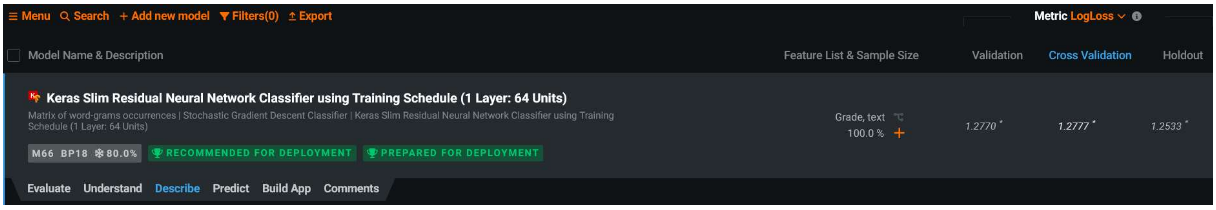


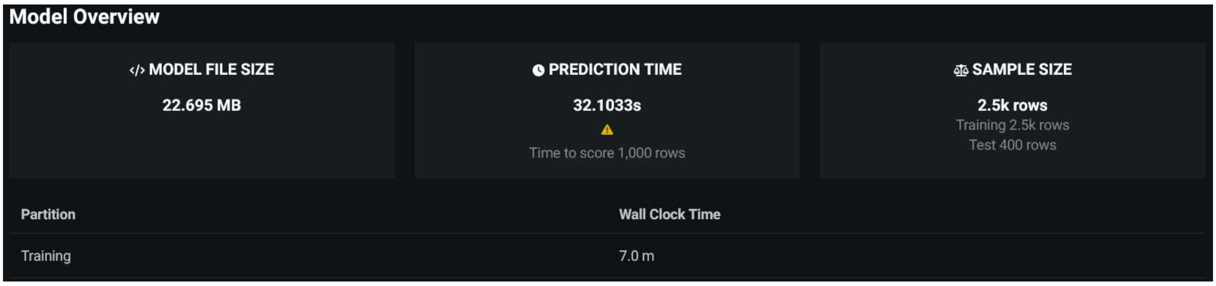


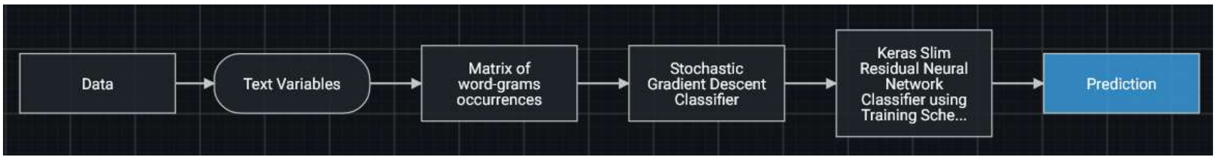
UniqueSTR36053H22.csv word clouds

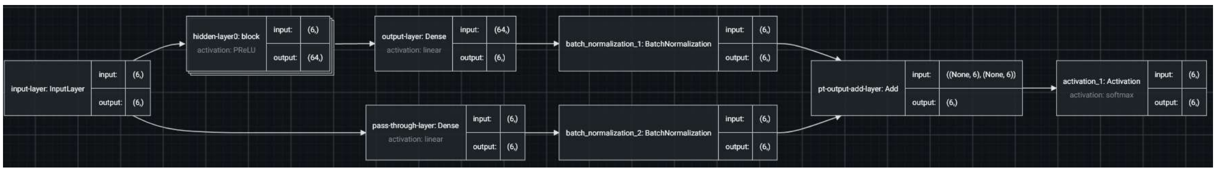
UniqueSTR36053all.csv data and model info







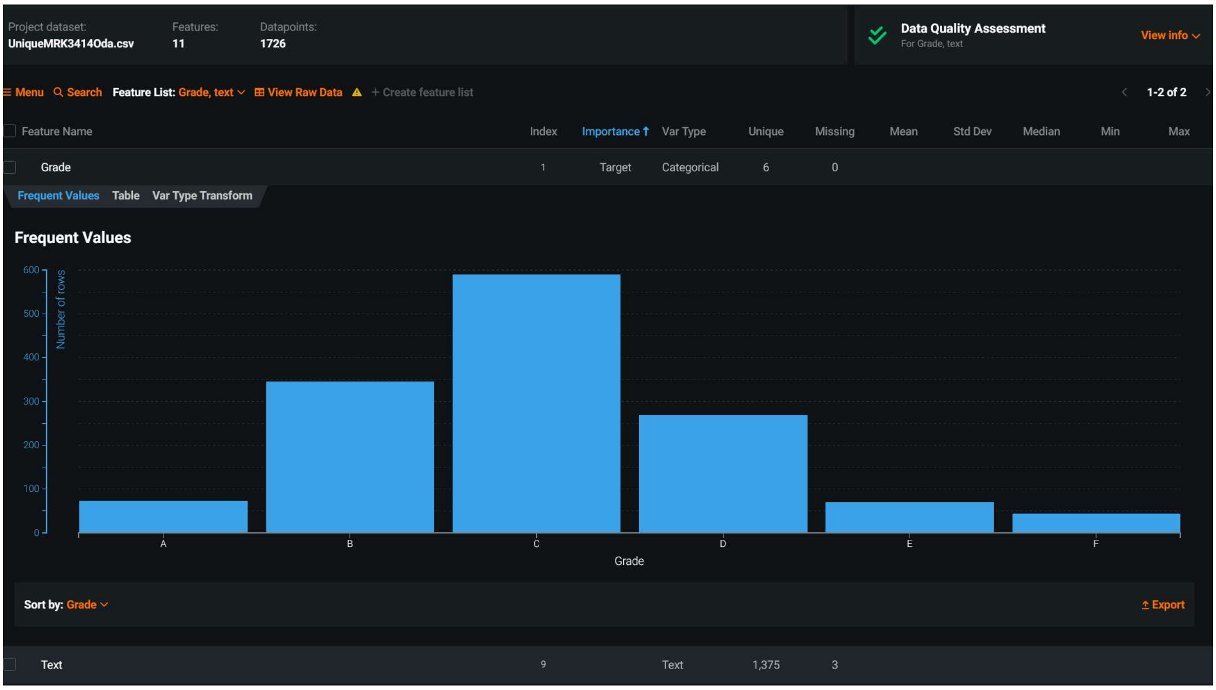


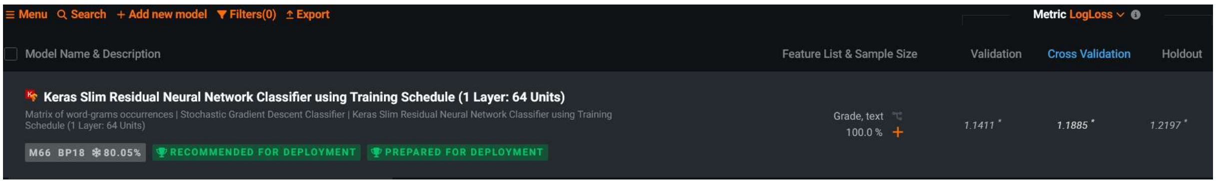


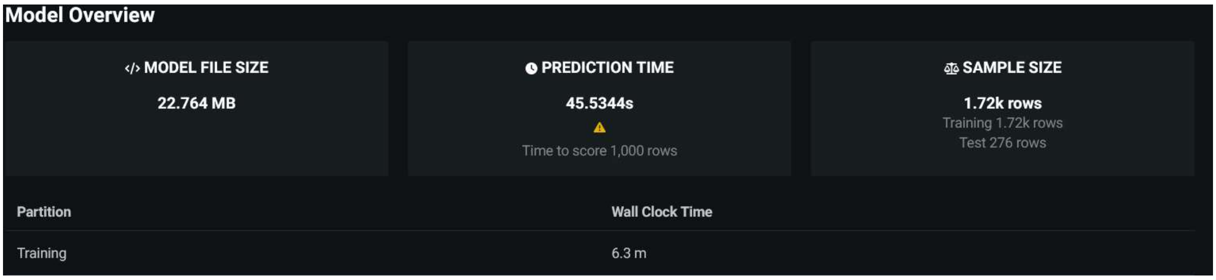
UniqueSTR36053all.csv confusion matrix

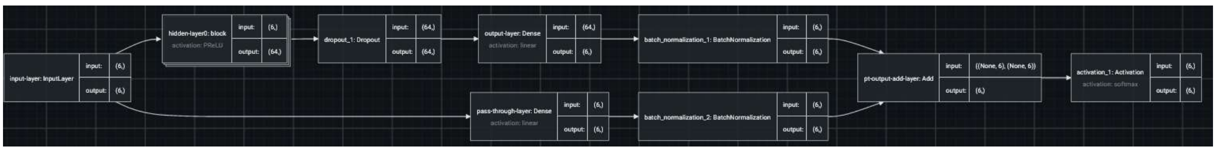
UniqueSTR36053all.csv word clouds

UniqueMRK3414Oda.csv data and model info

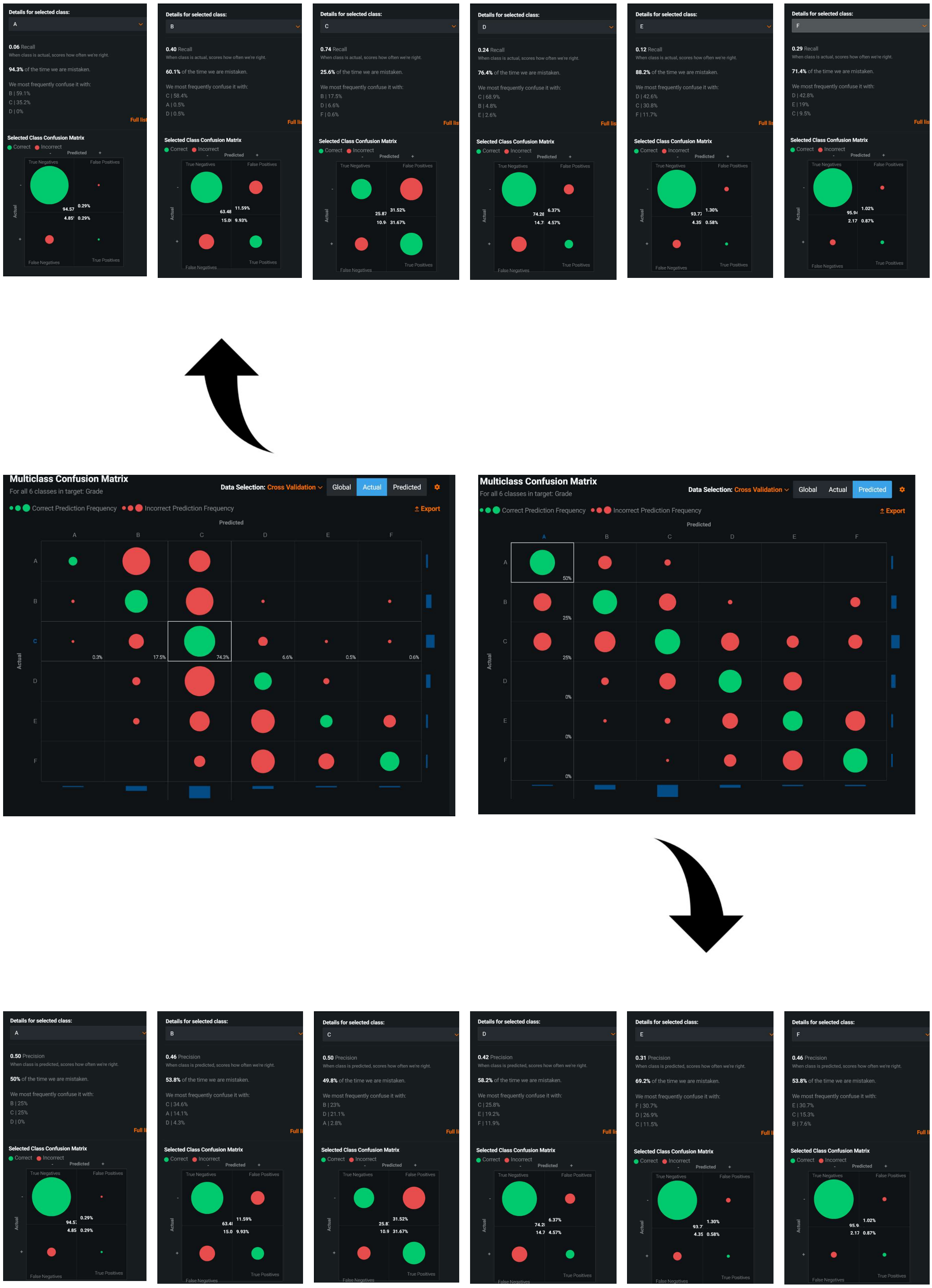






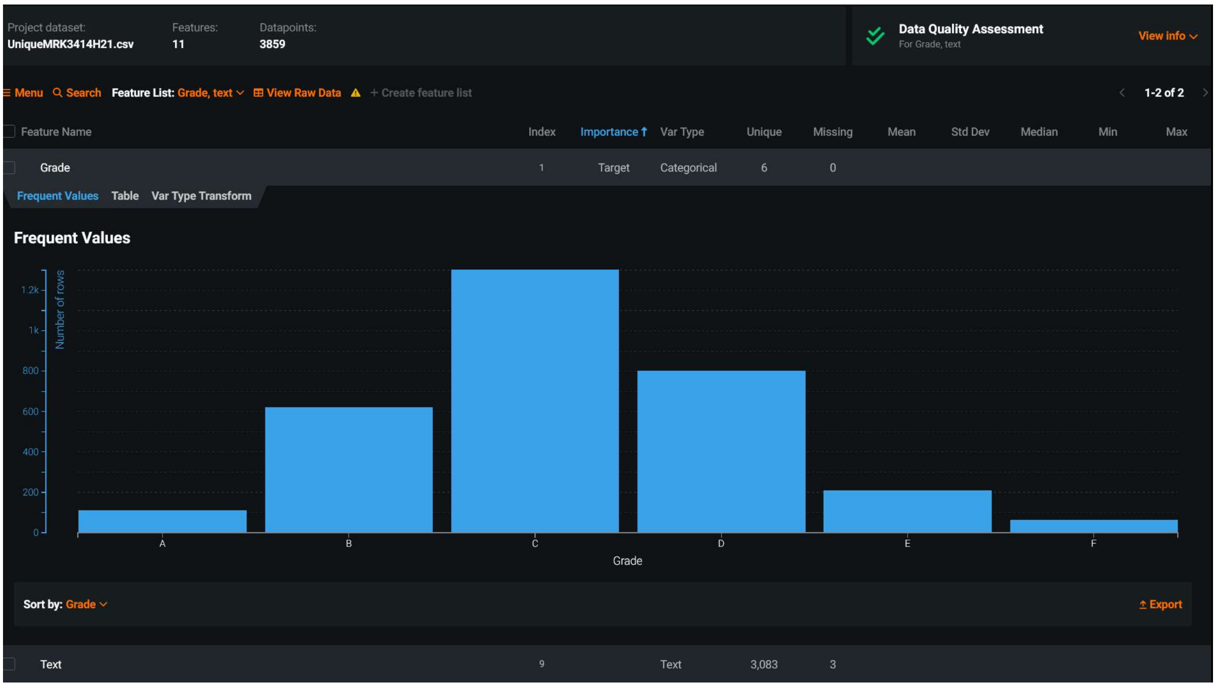


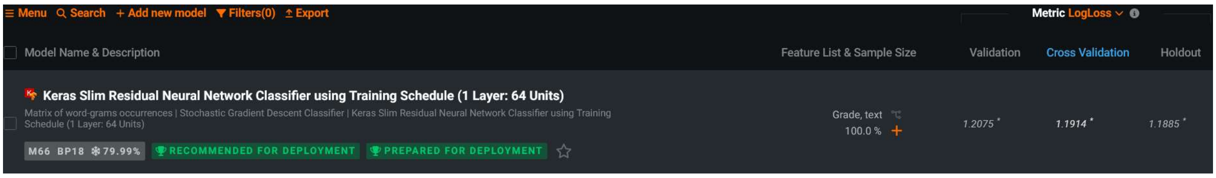
UniqueMRK3414Oda.csv confusion matrix

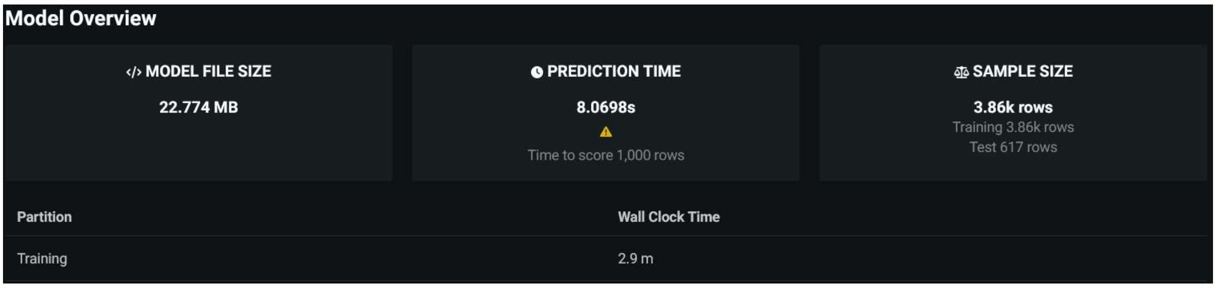


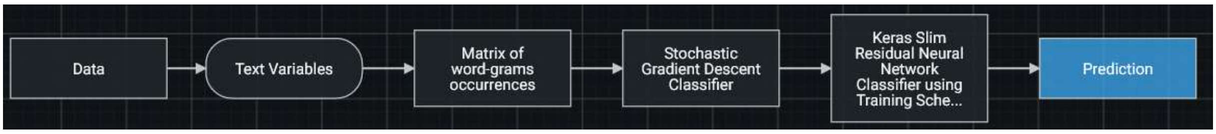
UniqueMRK3414Oda.csv word clouds

UniqueMRK3414H21.csv data and model info

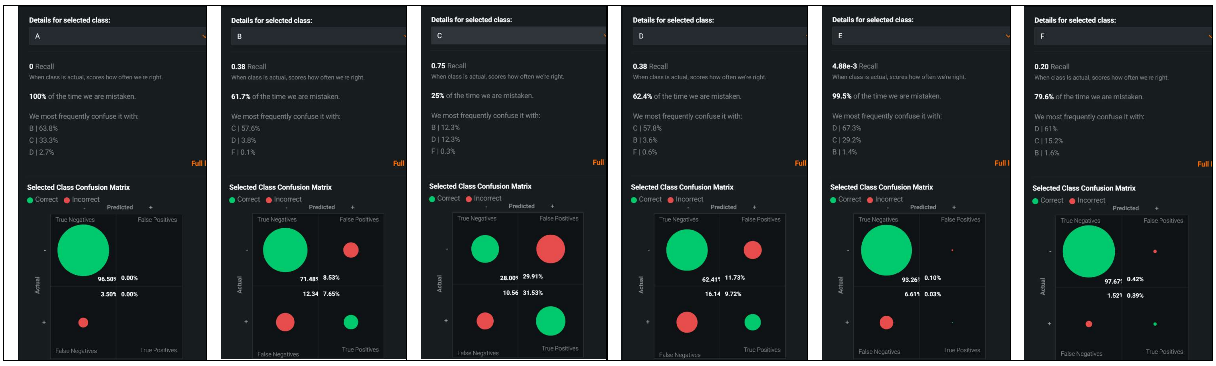


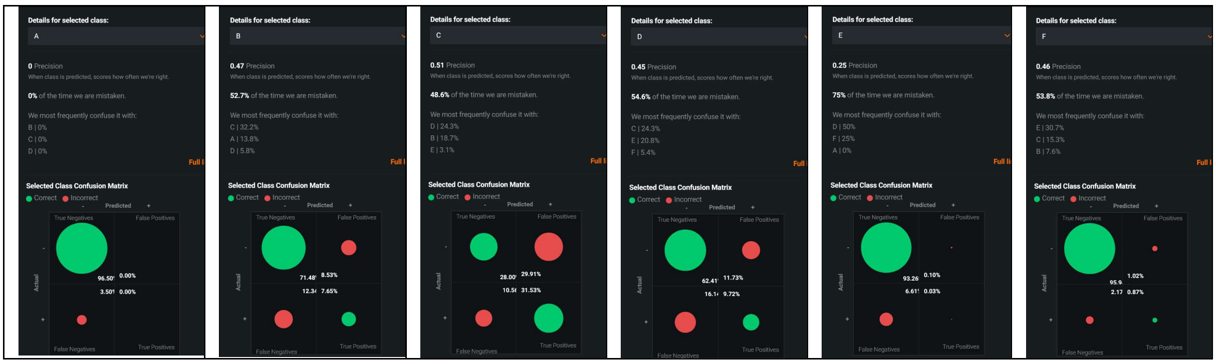






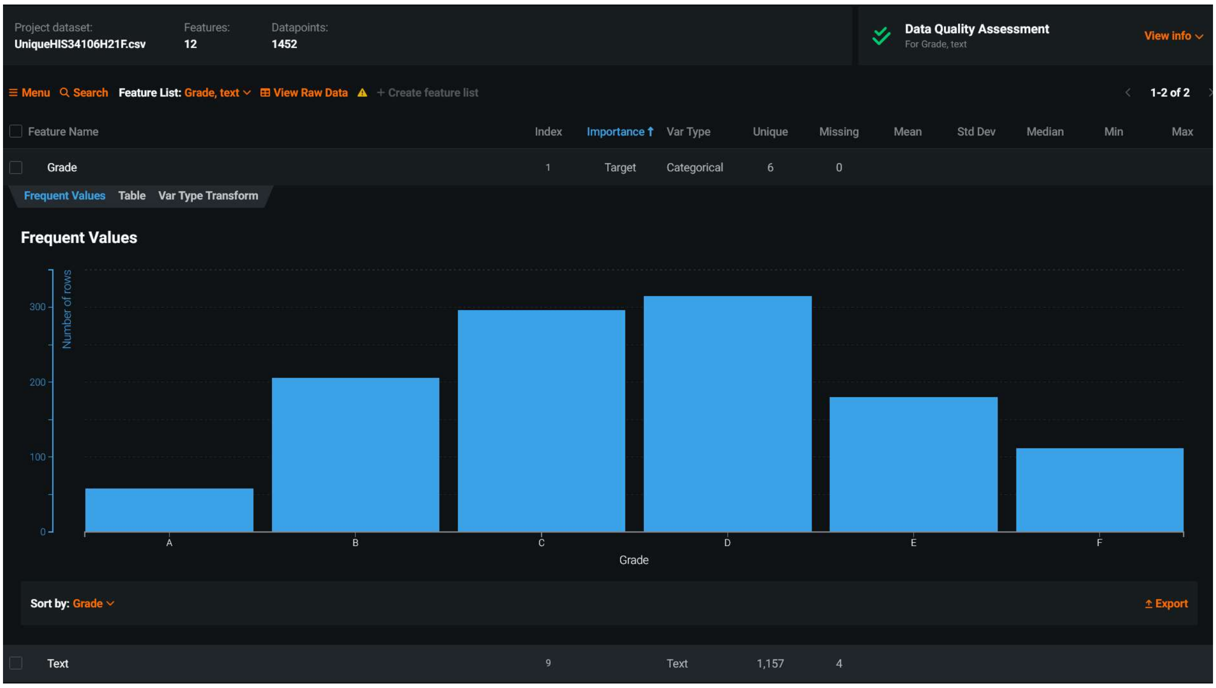
UniqueMRK3414H21.csv confusion matrix

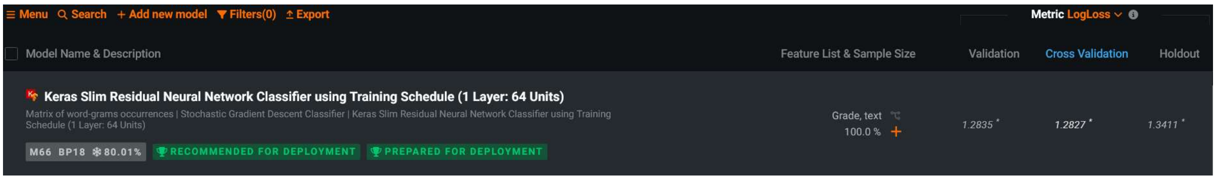


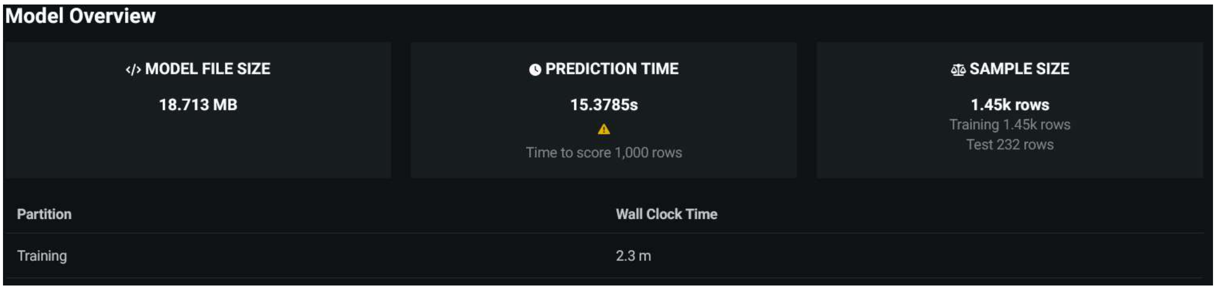


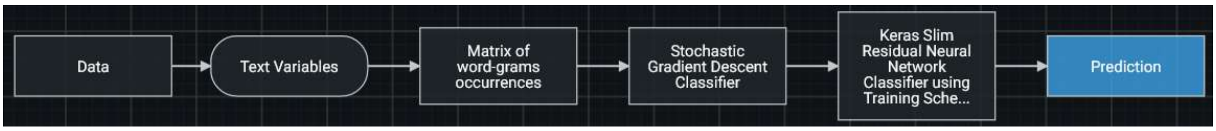
UniqueMRK3414H21.csv word clouds

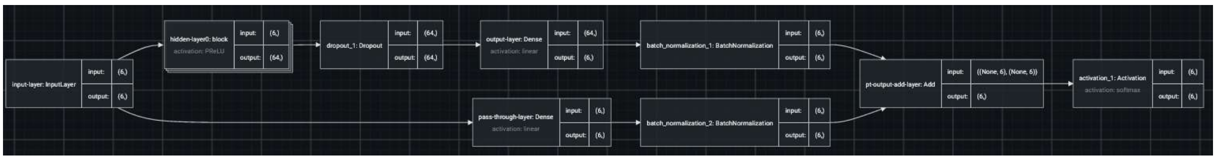
UniqueHIS34106H21F.csv data and model info



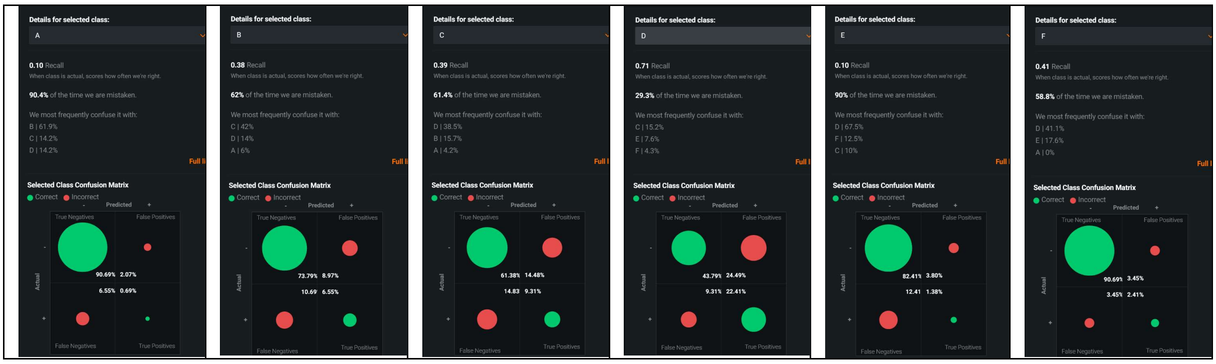


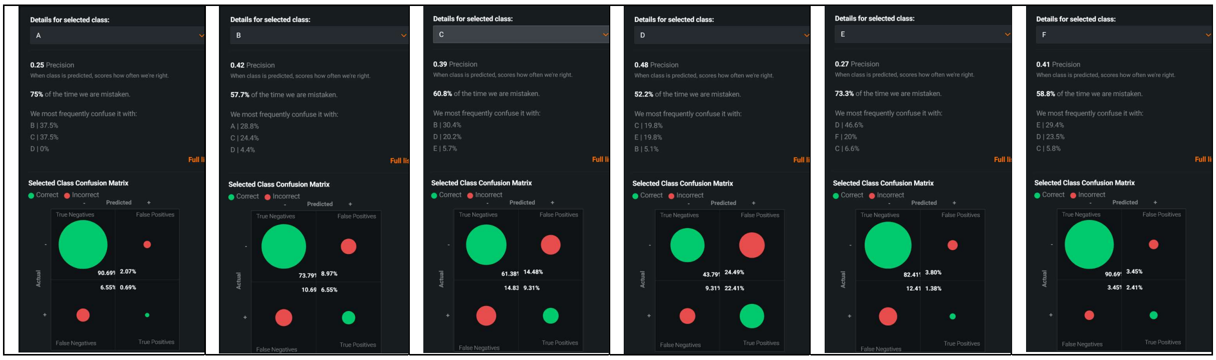






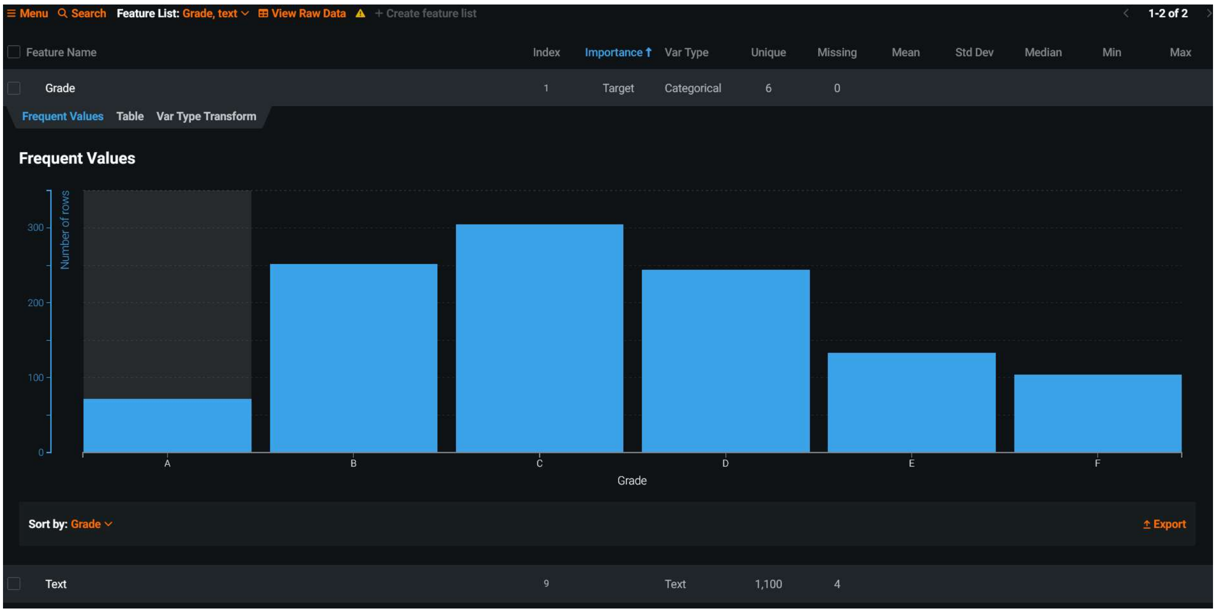
UniqueHIS34106H21F.csv confusion matrix

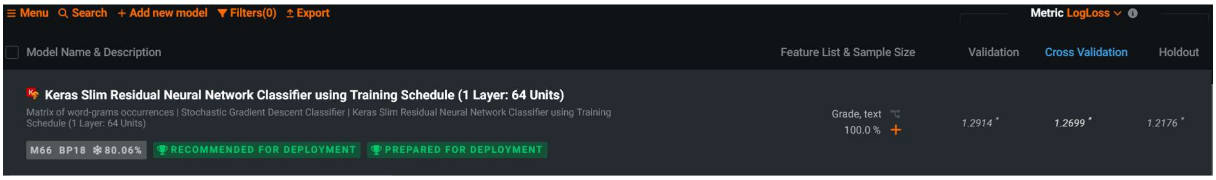


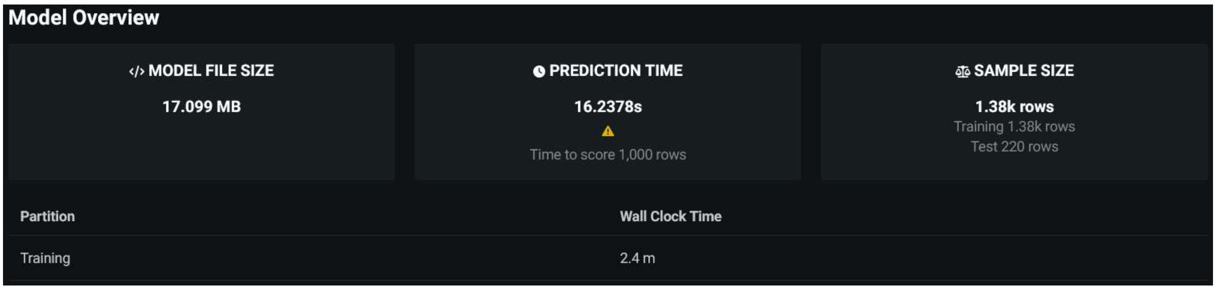


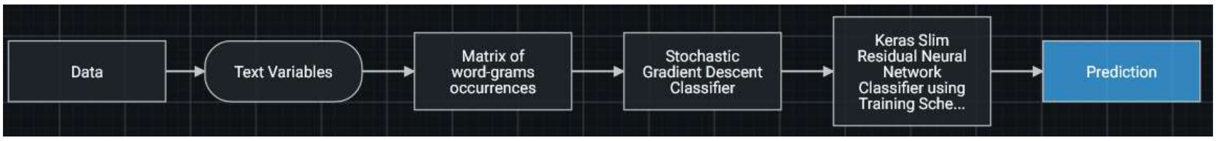
UniqueHIS34106H21F.csv word clouds

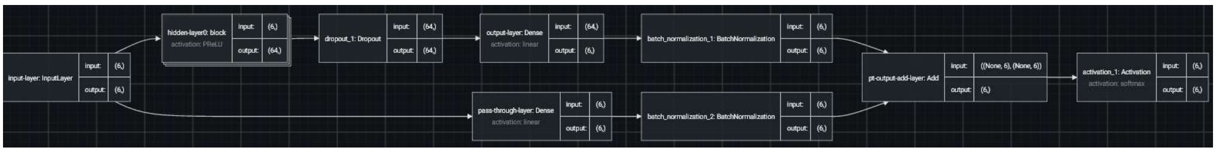
UniqueHIS34106H21E.csv data and model info



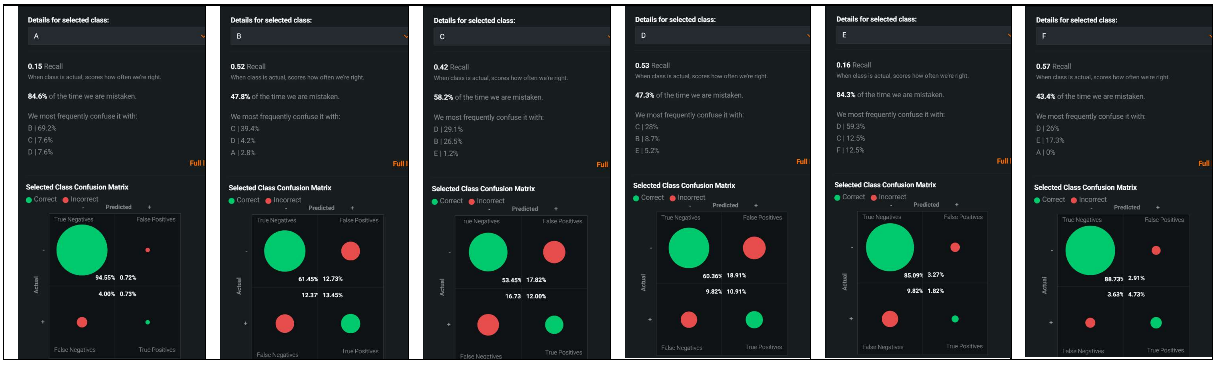


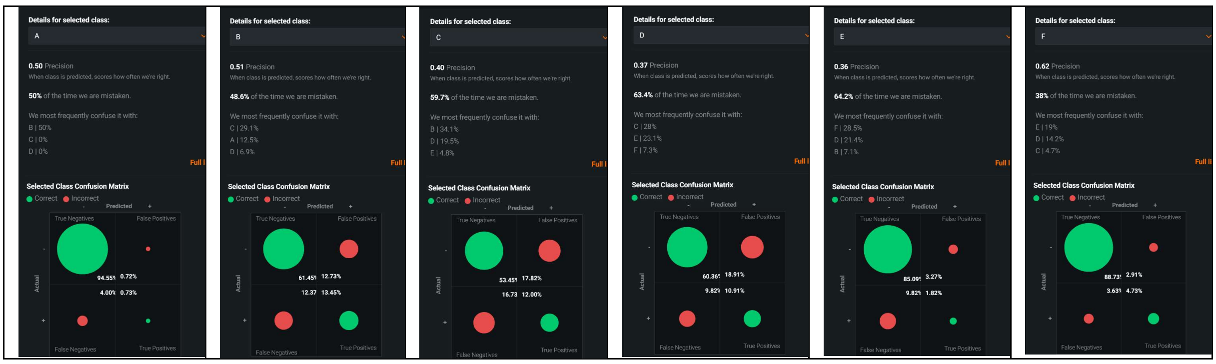






UniqueHIS34106H21E.csv confusion matrix





UniqueHIS34106H21E.csv word clouds

APPENDIX B

This notebook is for logging the work we're doing for our BI Automated Grading analytics project. Work was done in Spyder 3.9 and DataRobot.

There were many iterations to complete the code below. We have not include a detailed description of the iterations to produce a working code, but some warnings and main problems we faced below:

- nested in the participants column is a varying number of columns. We adapted the code for each of the jsons we processed with the code. If there are many jsons to process perhaps it's worth to improve the code to handle a varying number of columns there.
- the ordering of the columns also varied
- not all of the "unpacked" rows with grade and link to exam submission pdf had a link, but also not all links worked. These problems were solved.

The complete code below would not be neccessary to produce the datasets we ultimately used, but it was neccessary to properly explore the data.

The about csv which was exported in addition to the csv for use in the analyzis was only meant for creating an overview of the data.

Running the codes took between 5 and 30 minutes depending on the number and size of sumbissions.

We were running 12 kernels at the same time to process the jsons

In [ ]:
```
# Select file to work with
json = 6639770

# Import json file
import pandas as pd
df = pd.read_json(f'{json}.json', lines=True)


# Normalize and transpose to transform nested dictionary items into df.
flowData = pd.json_normalize(df.flowData)
assignment = pd.json_normalize(df.assignment)
Participants = pd.json_normalize(df.participants)
ParticipantsT = Participants.transpose()
ParticipantsT.columns = ['p']
ParticipantsT = pd.json_normalize(ParticipantsT.p)
ParticipantsT.columns = ['A', 'B', 'Grade', 'Comments', 'C', 'Handindate', 'D', 'T
Comments = pd.json_normalize(ParticipantsT.Comments)
Comments.columns = ['A', 'B', 'C', 'D']
Comment = pd.json_normalize(Comments.A)
Annotations = pd.json_normalize(ParticipantsT.C)
Annotations.columns = ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L',
Annotation = pd.json_normalize(Annotations.A)


# Concat About
```

```
About = pd.concat([assignment, flowData], axis=1)


# Concat relevant data. Leaving columns like IP adress, student ID out for privacy
rd = pd.concat([ParticipantsT.Grade, Comment.text, Comment.userId, Comment.timestam
rd = rd.rename(columns={rd.columns[1]: "Comment"})

#delete temporary dataframes
del df, assignment, flowData, Participants, ParticipantsT, Comment, Comments, Anno


# Decode HTML entities in rd to make Comment text readable
import html
import re


def clean_text(text):
    # Decode HTML entities
    decoded_text = html.unescape(text)

    # Remove HTML tags
    clean_text = re.sub('<.*?>', '', decoded_text)

    return clean_text

# Apply the function to every value in the 'comment' column
# Some rows were not strings, so changing all rows to strings first
rd['Comment'] = rd['Comment'].astype(str)
rd['Comment'] = rd['Comment'].apply(clean_text)



# Drop all rows with NaN value in the link column, NaN values there caused a proble
rd = rd.dropna(subset=['Link'], how='all')


# apply the function to each row in the dataframe and add the text to a new column
import requests
import io
from pypdf import PdfReader

# function to extract text from a PDF given a link
def extract_text_from_pdf(link):
    response = requests.get(link)
    try:
        with io.BytesIO(response.content) as data:
            pdf = PdfReader(data, strict=False)
            text = ''
            for page in pdf.pages:
                text += page.extract_text()
        return text
    except Exception:
        print(f"Error reading PDF at link {link}")
        return ""

# apply the function to each row in the dataframe and add the text to a new column
rd['Text'] = rd['Link'].apply(extract_text_from_pdf)


# save rd to a new csv file
rd.to_csv(f'Analyze{json}.csv', index=False)

# Save about to a new csv file
About.to_csv(f'About{json}.csv', index=False)
```

This is an overview of the exams we've recieved data on. The list is manually compiled from the downloaded about.csv's using spreadsheets\

Json	assignment.name	assignment.size	assignment.downloadUrl	title	subtitle	code	flowType
6639775	STR36053 Strategi - oppgave høst 2022.pdf	252628	https://api.eu.wiseflow.io/link-signer/v1/static/signed/4594e366b5ea5da60aa66ff4990259983fda410c870fdeee8a377502dc98c9d9/files/624382a7-ec1c-442b-a200-beaede957db9/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2022.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - R	202220||20715||IN08||R||P	FLOWassign
6639778	STR36053 Strategi - oppgave høst 2022.pdf	252628	https://api.eu.wiseflow.io/link-signer/v1/static/signed/769579d2d4b178bb08e2fab811d7b1d65b12e518b0a87b56433b2750ba24eed1/files/4fd8da26-c5e7-43e8-a5a2-a9fde4c9e8d3/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2022.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - T	202220||20617||IN08||T||P	FLOWassign
6639770	STR36053 Strategi - oppgave høst 2022.pdf	252628	https://api.eu.wiseflow.io/link-signer/v1/static/signed/23f426583aaa48f9ae0bc175ab6f2537b129b6af089b6e655e399d56dc2b6815/files/1b540c69-6af6-4ffc-b351-cffdb7e78327/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2022.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - F	202220||20491||IN08||F||P	FLOWassign
6639766	STR36053 Strategi - oppgave høst 2022.pdf	252628	https://api.eu.wiseflow.io/link-signer/v1/static/signed/546ed9943c496bfd5fad2b022a5f8f35a270ecbc3a6cbefe276edc3963ae5717/files/6ea1fb91-7f91-4aa7-8202-15bc5d919ed6/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2022.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - W	202220||20423||IN08||W||P	FLOWassign
6639761	STR36053 Strategi - oppgave høst 2022.pdf	252628	https://api.eu.wiseflow.io/link-signer/v1/static/signed/013639735e0bf01501f38abc29e4b84ff7eb9a70765ebcdf428573cd4bd64008/files/aebacf22-5547-4c4d-af41-a35965f89de2/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2022.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - B	202220||20840||IN08||B||P	FLOWassign
6450101	MRK3414Skriftlig eksamen h 2021.pdf	215131	https://api.eu.wiseflow.io/link-signer/v1/static/signed/4e6df04375db07d94d3a63b6caa53da814667c4eb3d7331c51b63cdbd28b0d70/files/fe8666f1-4a2e-4e65-a920-b0bbe3d6d9db/file?filename=MRK3414Skriftlig+eksamen+h+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	MRK 34143 Marked	Take-home examination 30% - F	202120||22470||IN11||F||H	FLOWassign
6450099	MRK3414Skriftlig eksamen h 2021.pdf	215131	https://api.eu.wiseflow.io/link-signer/v1/static/signed/4e6df04375db07d94d3a63b6caa53da814667c4eb3d7331c51b63cdbd28b0d70/files/fe8666f1-4a2e-4e65-a920-b0bbe3d6d9db/file?filename=MRK3414Skriftlig+eksamen+h+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	MRK 34143 Marked	Take-home examination 30% - B	202120||22366||IN11||B||H	FLOWassign
6450097	MRK3414Skriftlig eksamen h 2021.pdf	215131	https://api.eu.wiseflow.io/link-signer/v1/static/signed/4e6df04375db07d94d3a63b6caa53da814667c4eb3d7331c51b63cdbd28b0d70/files/fe8666f1-4a2e-4e65-a920-b0bbe3d6d9db/file?filename=MRK3414Skriftlig+eksamen+h+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	MRK 34143 Marked	Take-home examination 30% - R	202120||22365||IN11||R||H	FLOWassign
6450095	MRK3414Skriftlig eksamen h 2021.pdf	215131	https://api.eu.wiseflow.io/link-signer/v1/static/signed/4e6df04375db07d94d3a63b6caa53da814667c4eb3d7331c51b63cdbd28b0d70/files/fe8666f1-4a2e-4e65-a920-b0bbe3d6d9db/file?filename=MRK3414Skriftlig+eksamen+h+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	MRK 34143 Marked	Take-home examination 30% - T	202120||22364||IN11||T||H	FLOWassign
6450093	MRK3414Skriftlig eksamen h 2021.pdf	215131	https://api.eu.wiseflow.io/link-signer/v1/static/signed/4e6df04375db07d94d3a63b6caa53da814667c4eb3d7331c51b63cdbd28b0d70/files/fe8666f1-4a2e-4e65-a920-b0bbe3d6d9db/file?filename=MRK3414Skriftlig+eksamen+h+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	MRK 34143 Marked	Take-home examination 30% - W	202120||22363||IN11||W||H	FLOWassign
6445370	HIS 34106_202120_01.12.2021_14.00-17.00	119495	https://api.eu.wiseflow.io/link-signer/v1/static/signed/6577ac0df095cc7a3682fe413f1e2c82a5c00e62ff1d6b66eb9f2880d8631eac/files/a3ae7c7c-cef2-4061-8e3e-feceee714f42/file?filename=HIS+34106_202120_01.12.2021_14.00-17.00_QP.pdf&content_type=application%2Fpdf&content_disposition=attachment	HIS 34106 Bedriften	Written examination 100% - T	202120||22622||IN02||T||E	FLOWlock
6445368	HIS 34106_202120_01.12.2021_14.00-17.00	119495	https://api.eu.wiseflow.io/link-signer/v1/static/signed/d7d313c3539a1d4dacb948ac1c01043817165e3c4faaf83d66a9de1a9227c36d/files/bb9ebddd-252e-44a7-8f88-77eda0149acc/file?filename=HIS+34106_202120_01.12.2021_14.00-17.00_QP.pdf&content_type=application%2Fpdf&content_disposition=attachment	HIS 34106 Bedriften	Written examination 100% - R	202120||22621||IN02||R||E	FLOWlock
6445365	HIS 34106_202120_01.12.2021_14.00-17.00	119495	https://api.eu.wiseflow.io/link-signer/v1/static/signed/747cdcfae98b57d29a438fe19a5d9fe66d2fe3c33fe40e687aedb923df60fba3/files/630a91a8-adc8-4196-8fd8-6eab0fead203/file?filename=HIS+34106_202120_01.12.2021_14.00-17.00_QP.pdf&content_type=application%2Fpdf&content_disposition=attachment	HIS 34106 Bedriften	Written examination 100% - W	202120||22620||IN02||W||E	FLOWlock
6445354	HIS 34106_202120_01.12.2021_09.00-12.00	101751	https://api.eu.wiseflow.io/link-signer/v1/static/signed/ade6b690f2cbf8dc0e1d056f1dccf8faced6cd24568e170d511a8a88ed90a1f7/files/6780814d-e393-4e03-b56c-6a8f854c0b98/file?filename=HIS+34106_202120_01.12.2021_09.00-12.00_QP.pdf&content_type=application%2Fpdf&content_disposition=attachment	HIS 34106 Bedriften	Written examination 100% - B	202120||20917||IN02||B||E	FLOWlock
6445350	HIS 34106_202120_01.12.2021_09.00-12.00	101751	https://api.eu.wiseflow.io/link-signer/v1/static/signed/e730529ba03bea3adaa52434cc6d09cff334c10ae61df468977f798df43333fc/files/2dd8f2b2-ecb8-432a-8a0f-7d167563b8fa/file?filename=HIS+34106_202120_01.12.2021_09.00-12.00_QP.pdf&content_type=application%2Fpdf&content_disposition=attachment	HIS 34106 Bedriften	Written examination 100% - R	202120||20804||IN02||R||E	FLOWlock
6445346	HIS 34106_202120_01.12.2021_09.00-12.00	101751	https://api.eu.wiseflow.io/link-signer/v1/static/signed/9bd4b0915cabde982d3b47d509be92f5c238f9f3bdb8194e39765bad3ba9a41e/files/0e50708f-c2e5-41bd-90a6-f5545822085f/file?filename=HIS+34106_202120_01.12.2021_09.00-12.00_QP.pdf&content_type=application%2Fpdf&content_disposition=attachment	HIS 34106 Bedriften	Written examination 100% - T	202120||20735||IN02||T||E	FLOWlock
6445342	HIS 34106_202120_01.12.2021_09.00-12.00	101751	https://api.eu.wiseflow.io/link-signer/v1/static/signed/200de194cec1433231141591032f3d6582b271bc7c9ae8f2344a841179168f5f/files/1426c190-88ae-4a2c-ad1e-de7d23f0c6f3/file?filename=HIS+34106_202120_01.12.2021_09.00-12.00_QP.pdf&content_type=application%2Fpdf&content_disposition=attachment	HIS 34106 Bedriften	Written examination 100% - W	202120||20413||IN02||W||E	FLOWlock
6386998	MRK 3414 Markedsføringsledelse Caseteks	1019908	https://api.eu.wiseflow.io/link-signer/v1/static/signed/3ac2cb959925fcb74f185de8e0835a509dbdb117ca278c6ecadfa44a8c081f53/files/c8640d04-15e8-4c8e-af22-d9e043174135/file?filename=MRK+3414+Markedsf%C3%B8ringsledelse+Casetekst+Oda+Konte.pdf&content_type=application%2Fpdf&content_dispositi	MRK 34141 Marked	Term paper 100% - W	202120||22068||IN11||W||P	FLOWassign
6346624	MRK 3414 Markedsføringsledelse Caseteks	1016150	https://api.eu.wiseflow.io/link-signer/v1/static/signed/401dd5a67179ffcfa2e5946dad2ca9027578a18670c64211c8aafb1f3d5c994f/files/20c082cb-6ab6-4651-a23e-dd00cc0d43df/file?filename=MRK+3414+Markedsf%C3%B8ringsledelse+Casetekst+Oda+%28002%29.pdf&content_type=application%2Fpdf&content_dis	MRK 34142 Marked	Term paper 70% - R	202120||20809||IN11||R||P	FLOWassign
6346603	MRK 3414 Markedsføringsledelse Caseteks	1016150	https://api.eu.wiseflow.io/link-signer/v1/static/signed/052b47896377c7eadb8ae8388102104c55987ad197e9e73c6de433d7cc37d992/files/0fb4501c-0729-45f1-894d-bfe9cb0ea4e0/file?filename=MRK+3414+Markedsf%C3%B8ringsledelse+Casetekst+Oda+%28002%29.pdf&content_type=application%2Fpdf&content_	MRK 34142 Marked	Term paper 70% - T	202120||20741||IN11||T||P	FLOWassign
6346582	MRK 3414 Markedsføringsledelse Caseteks	1016150	https://api.eu.wiseflow.io/link-signer/v1/static/signed/dbbdf5a477c6794ba68a764d1be26dcc72fe6202fc6e5386308774707f7dffca/files/6cb8ad52-925f-44d2-8efe-5867fbbc88da/file?filename=MRK+3414+Markedsf%C3%B8ringsledelse+Casetekst+Oda+%28002%29.pdf&content_type=application%2Fpdf&content_disp	MRK 34142 Marked	Term paper 70% - F	202120||20643||IN11||F||P	FLOWassign
6345085	STR36053 Strategi - oppgave høst 2021.pdf	142874	https://api.eu.wiseflow.io/link-signer/v1/static/signed/1a72ea018523607a17cb6591b56cec1dc064eef414bcb2ca01913b5d69c3f333/files/62ab263d-c4d4-454f-b6f4-e195f563fa18/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - B	202120||20935||IN08||B||P	FLOWassign
6345079	STR36053 Strategi - oppgave høst 2021.pdf	142874	https://api.eu.wiseflow.io/link-signer/v1/static/signed/1a72ea018523607a17cb6591b56cec1dc064eef414bcb2ca01913b5d69c3f333/files/62ab263d-c4d4-454f-b6f4-e195f563fa18/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - R	202120||20821||IN08||R||P	FLOWassign
6345067	STR36053 Strategi - oppgave høst 2021.pdf	142874	https://api.eu.wiseflow.io/link-signer/v1/static/signed/1a72ea018523607a17cb6591b56cec1dc064eef414bcb2ca01913b5d69c3f333/files/62ab263d-c4d4-454f-b6f4-e195f563fa18/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - T	202120||20753||IN08||T||P	FLOWassign
6345058	STR36053 Strategi - oppgave høst 2021.pdf	142874	https://api.eu.wiseflow.io/link-signer/v1/static/signed/1a72ea018523607a17cb6591b56cec1dc064eef414bcb2ca01913b5d69c3f333/files/62ab263d-c4d4-454f-b6f4-e195f563fa18/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - F	202120||20650||IN08||F||P	FLOWassign
6331885	STR36053 Strategi - oppgave høst 2021.pdf	142874	https://api.eu.wiseflow.io/link-signer/v1/static/signed/1a72ea018523607a17cb6591b56cec1dc064eef414bcb2ca01913b5d69c3f333/files/62ab263d-c4d4-454f-b6f4-e195f563fa18/file?filename=STR36053+Strategi+-+oppgave+h%C3%B8st+2021.pdf&content_type=application%2Fpdf&content_disposition=attachment	STR 36053 Strategi	Term paper 100% - W	202120||20478||IN08||W||P	FLOWassign
6331234	MRK 3414 Markedsføringsledelse Caseteks	1016150	https://api.eu.wiseflow.io/link-signer/v1/static/signed/76c2cf3b52df051d70e0a4a2e76c80821d1877f14c07541531565339a7a77348/files/9891fef6-4b15-4c55-9833-5042099d16dd/file?filename=MRK+3414+Markedsf%C3%B8ringsledelse+Casetekst+Oda+%28002%29.pdf&content_type=application%2Fpdf&content_di	MRK 34142 Marked	Term paper 70% - W	202120||20441||IN11||W||P	FLOWassign
6346657							

Next step should have been to download the assignment text to verify which csv's with exam data we could combine. Obviously we needed to keep submissions to different assignments separate and only train models where all submissions are answering the same assignment.

Instead we looked at assignment file size to identify which assignments were the same. This led to a mistake. The STR36053H21 and STR36053H22 had identically worded assignments, but had different file sizes. So we ended up separating them into two datasets, one for each year of that course.

When we finally downloaded the actual assignement pdf's later, we did not use code to download those pdf,s as there were only 6 (5) of them. We did not download for each json, just each json belonging to a spesific course code. We ran the code normalizing the assignment column for one of the jsons from each course code and manually copied the url link from the dataframe explorer in spyder.

After verifying which jsons to concatenate, next step was to concatenate the appropriate data. We had one json for each campus that had students submitting exams to the courses we were investigating.

We ran this code for each of the unique courses we'd recieved jsons on. There were some adaptations neccessary for each time we ran the code:

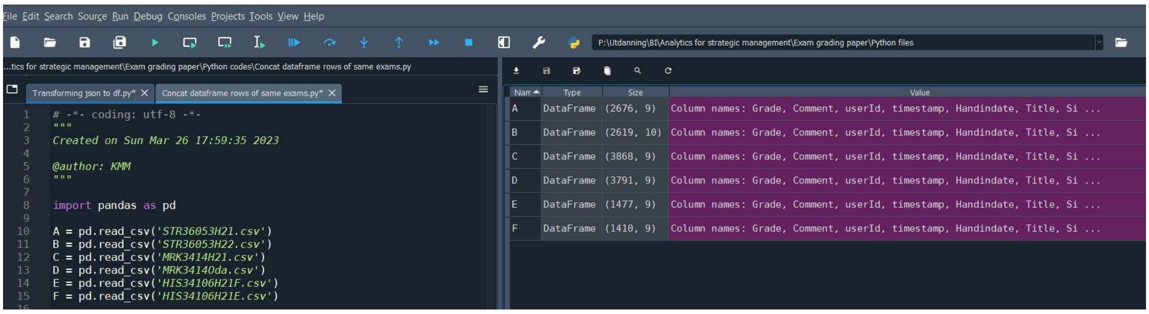
- The number of jsons for each course varied between 5-7.
- Which jsons that had missings columns that needed to be inserted before concatenating varied

In [ ]:

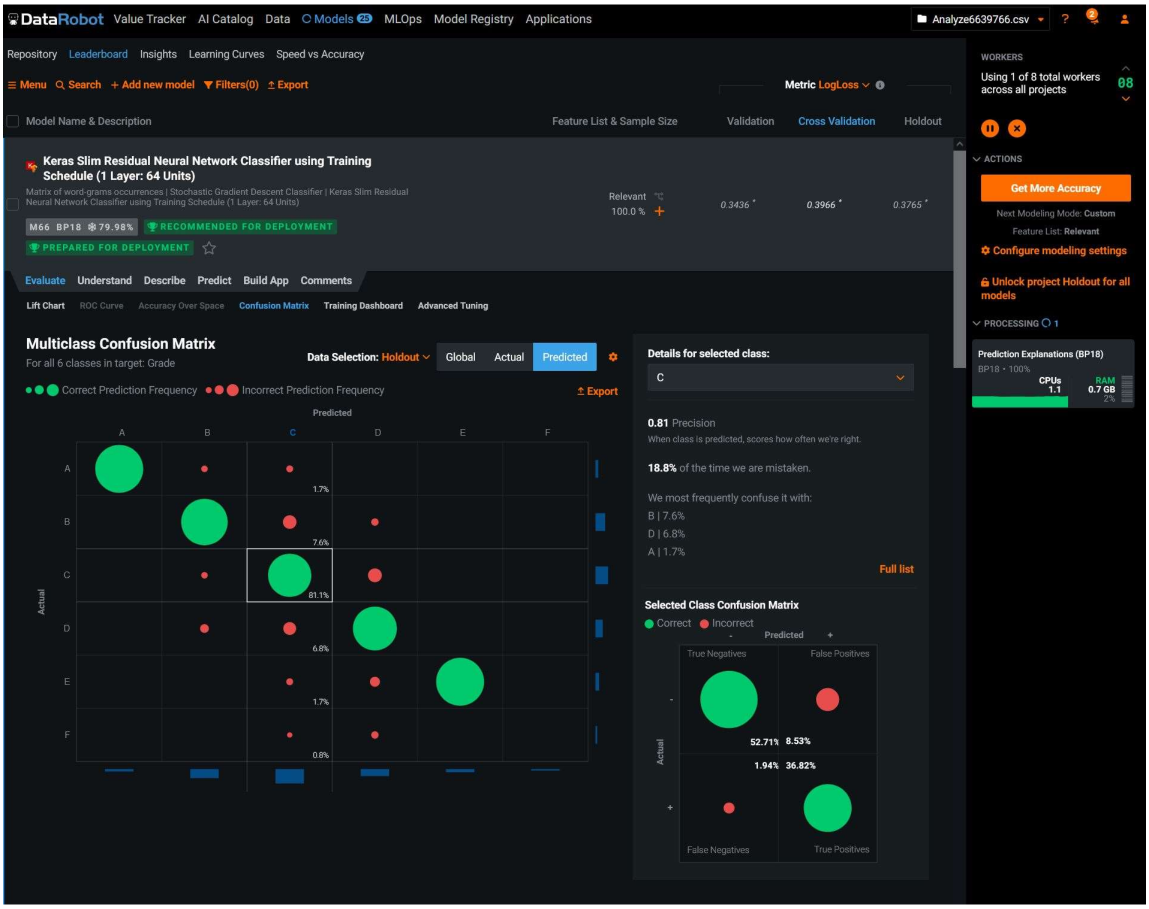
```
import pandas as pd

A = pd.read_csv('Analyze6639775.csv')
B = pd.read_csv('Analyze6639778.csv')
C = pd.read_csv('Analyze6639770.csv')
D = pd.read_csv('Analyze6639766.csv')
E = pd.read_csv('Analyze6639761.csv')


#Adding missing columns in some dataframes
C.insert(2, 'userId', '')
C.insert(3, 'timestamp', '')

D.insert(2, 'userId', '')
D.insert(3, 'timestamp', '')
```

```
# concatenate the dataframes into a single dataframe
All = pd.concat([A, B, C, D, E], ignore_index=True)

# save rd to a new csv file
All.to_csv('STR36053H22.csv', index=False)
```

Here is an overview of the rows we had for each exam:



Iteration one included testing model training on data from single jsons. We did not realize the source of the dataleak until jsons were concatenated as described above. Iteration 1 led to the discovery of a dataleak as described in the paper. We got 8/8 A's and 15/15 E's. Rest was also very good.



Here is the code used to remove duplicates from the different datasets.

In [ ]:
```
import pandas as pd

A = pd.read_csv('STR36053H21.csv')
B = pd.read_csv('STR36053H22.csv')
C = pd.read_csv('MRK3414H21.csv')
```

```
D = pd.read_csv('MRK3414Oda.csv')
E = pd.read_csv('HIS34106H21F.csv')
F = pd.read_csv('HIS34106H21E.csv')

G = A.drop_duplicates(subset=['Grade', 'Text'])
H = B.drop_duplicates(subset=['Grade', 'Text'])
I = C.drop_duplicates(subset=['Grade', 'Text'])
J = D.drop_duplicates(subset=['Grade', 'Text'])
K = E.drop_duplicates(subset=['Grade', 'Text'])
L = F.drop_duplicates(subset=['Grade', 'Text'])

G.to_csv('UniqueSTR36053H21.csv', index=False)
H.to_csv('UniqueSTR36053H22.csv', index=False)
I.to_csv('UniqueMRK3414H21.csv', index=False)
J.to_csv('UniqueMRK3414Oda.csv', index=False)
K.to_csv('UniqueHIS34106H21F.csv', index=False)
L.to_csv('UniqueHIS34106H21E.csv', index=False)
```

Iteration 2 was training models in DataRobot as described in the paper.

In iteration 3 we wanted to test running predictions and therefore needed to create a "pre DataRobot" train test split. We used this code:

In [ ]:
```
import pandas as pd
from sklearn.model_selection import train_test_split

A = pd.read_csv('UniqueSTR36053H21.csv')

# Set the partition ratio
train_ratio = 0.8
test_ratio = 0.2

# Perform the partitioning using train_test_split
A_train, A_test = train_test_split(A, train_size=train_ratio, test_size=test_ratio

# save train and holdout splits to new csv files
A_train.to_csv('STR36053H21_train.csv', index=False)
A_test.to_csv('STR36053H21_holdout.csv', index=False)
```

In iteration 4 we wanted to explore the effects of increasing the quantity of data available to train on as described in the paper. We used this code to prepare the data for analysis:

In [ ]:
```
import pandas as pd

A = pd.read_csv('UniqueSTR36053H21.csv')
B = pd.read_csv('UniqueSTR36053H22.csv')

# concatenate the dataframes into a single dataframe
All = pd.concat([A, B], ignore_index=True)

# save concatenated to a new csv file
All.to_csv('UniqueSTR36053all.csv', index=False)
```

Iteration 5 did not involve any data preparation. Only retraining with a different feature selection as described in the paper which is done in datarobot.

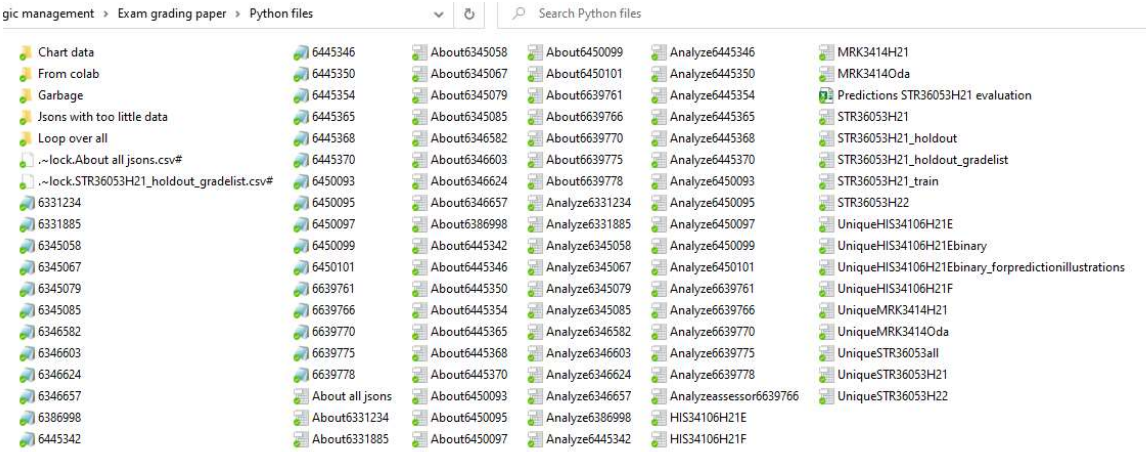
Iteration 6 involved creating a new column with binary values on whether the graded were A,B,C or D,E,F to test how a binary classification model would perform. We used the code below prepare the data for analysis.

In [ ]:
```
import pandas as pd

A = pd.read_csv('UniqueHIS34106H21E.csv')

# Function to categorize grades
def categorize_grade(Grade):
    if Grade in ['A', 'B', 'C']:
        return 1
    elif Grade in ['D', 'E', 'F']:
        return 0

# Create a new column 'Category' based on the 'Grade' column
A['Category'] = A['Grade'].apply(categorize_grade)

A.to_csv('UniqueHIS34106H21Ebinary.csv', index=False)
```

We fed 100% of our data to DataRobot, but for illustration purposes we went back to make a test train split on the same data in order to run a prediction. This is the code we used:

In [ ]:
```
# partition made for illustration purposes
from sklearn.model_selection import train_test_split

# Set the partition ratio
train_ratio = 0.95
test_ratio = 0.05

# Perform the partitioning using train_test_split
A_train, A_test = train_test_split(A, train_size=train_ratio, test_size=test_ratio

# save test split to a new csv file
A_test.to_csv('UniqueHIS34106H21Ebinary_forpredictionillustrations.csv', index=Fal
```

List of files used during the analysis:



Ved karaktersetting teller langsvar 70 prosent og kortsvar 30 prosent.
Du må bestå både langsvars- og kortsvarspørsmålene for å bestå eksamen.

Du bør bruke omlag 2 timer og 15 minutter på langsvarsoppgaven,
og omlag 45 minutter på de tre kortsvarspørsmålene.

Sensorene vet du har dårlig tid, så ikke mist motet selv om det blir hektisk, og du synes du ikke får prestert som du ønsker.

1. Langsvar: En-1- av oppgavene skal besvares (70%) – estimert tid 2 t 15 min

Enten 1A

Redegjør hva som menes med intern og ekstern frihet i pensum. Dernest, skal du redegjøre og forklare hvordan intern og ekstern frihet har variert fra bedriftens førindustrielle historie og gjennom de tre industrielle revolusjoner.

eller 1B

Redegjør for hva som menes med bedrifters eierskaps- og lokaliseringsfortrinn. Bruk disse to begrepene til å belyse hovedtrekk ved utviklingen av norsk oljenæring.

2. Kortsvar: Alle oppgavene skal besvares (30%) – estimert tid 45 min.

1. Hva menes med begrepet «kapitalisme»?
2. Forklar hvilken rolle ytringsklimaet i en bedrift kan spille for å forebygge normalisering av tvilsom adferd.
3. Gjør rede for Frederick W. Taylor og hans program «Scientific Management», også kalt taylorisme.

Ved karaktersetting teller langsvar 70 prosent og kortsvar 30 prosent.
Du må bestå både langsvars- og kortsvarspørsmålene for å bestå eksamen.

Du bør bruke omlag 2 timer og 15 minutter på langsvarsoppgaven,
og omlag 45 minutter på de tre kortsvarspørsmålene.

Sensorene vet du har dårlig tid, så ikke mist motet selv om det blir hektisk, og du synes du ikke får prestert som du ønsker.

1. Langsvar: En – 1 – av oppgavene skal besvares (70%) – estimert tid 2 t 15 min.

Enten 1A

Gjør rede for Frederick W. Taylor og hans program «Scientific Management», også kalt taylorisme. På hvilken måte kan de japanske ledelsesidealer, som vokste frem på 1970-tallet, sies å være en forlengelse av Taylors ideer, og på hvilken måte skilte de seg fra Taylor?

eller 1B

Redegjør for «egeninteressens problem» slik det er presentert pensum. Drøft deretter ulike svar eller løsninger på dette problemet.

2. Kortsvar: Alle oppgavene skal besvares (30%) – estimert tid 45 min.

1. Hva innebærer det at et aksjeselskap har begrenset ansvar?
2. Hva menes med paternalisme?
3. Hvilke momenter har blitt lansert for å forklare tendensen til *de*globalisering de siste årene?

MRK 34141 Markedsføringsledelse, høst 2021

Lengde på besvarelse: maks 26 sider. Vedlegg tillatt
Det skal utarbeides en innholdsfortegnelse og et sammendrag først i besvarelsen og en litteraturliste bakerst i besvarelsen. Disse sidene kommer i tillegg til selve eksamensbesvarelsen. En utførlig forklaring om sitering og referanseteknikk og mal for oppgaveskriving finner du på Student Portalen
https://portal.bi.no/eksamen-og-oppgave/oppgaveskriving/

FORMELLE KRAV – LES DETTE NØYE

Oppgaven skal løses individuelt eller i grupper på inntil tre (3) personer. Samarbeid mellom flere studenter /grupper om utarbeidelse av besvarelsen blir betraktet som fusk eller forsøk på fusk og rammes av forskrift om opptak, studier og eksamen for Handelshøyskolen BI, som det forutsettes at studentene er kjent med.

Gruppeinnlevering forutsetter eksamenspåmelding i identisk eksamenskode og ved samme eksamenssted. Det betyr samme campus. Oslo studentene som går i ulike klasser kan skrive i samme gruppe, men kan ikke ha gruppemedlem fra Bergen eller annen campus. Alle i Oslo kan skrive sammen. Alle i Bergen sammen etc. Nettstudentene er på egen CRN og kan ikke danne gruppe med de som går på heltidsstudiene, nettstudenter kan bare danne gruppe med nettstudenter.

Sideantall og rekkefølge av dokumenter:

Del 1: Markedsplanen; følg det oppsettet om er gitt i læreboken (Sammendrag, Markedsoversikt, Markedsmål, Markedsstrategi, Økonomi og Kontroll). Denne delen er maksimum 10 sider. I denne delen skal det ikke være referanseliste og vedlegg. Bruk av fotnoter er tillatt, men omfanget bør begrenses (maksimum fem fotnoter i markedsplanen).

Del 2: Faglig begrunnelse; følg oppsettet gitt senere i dette dokumentet (*"Analyse av bedriftens strategiske markedsutfordringer"* og *"Bedriftens markedsstrategiske tiltak").* Avsluttes med referanseliste (kildehenvisninger). Denne delen er maksimum 15 sider pluss vedlegg og referanseliste. Det anbefales at ca 2/3 av faglig begrunnelse benyttes til analysedelen og 1/3 til bedriftens markedsstrategiske tiltak.

Del 3: Vedlegg (utdrag fra øvelsesoppgavene og annen dokumentasjon - maks 20 sider vedlegg).

Del 4: Egenevaluering.

De fire delene skal være i ett dokument. Det er tillatt å ha skilleark (med overskriftstekst) mellom delene (skilleark teller ikke som sidetall).

Prosjektoppgaven i MRK 3414 Markedsføringsledelse går ut på å utarbeide en markedsplan for bedriften Oda i det norske markedet. Dere skal også gi en faglig begrunnelse for markedsplanen. Oda er nærmere beskrevet i vedlagte casetekst.

Det er ikke anledning til å intervjue ansatte hos Oda eller noen annen bedrift i forbindelse med løsning av prosjektoppgaven. Dette fordi det er et meget stort antall studenter som løser oppgaven og for den enkelte bedrift vil det kunne bli en urimelig belastning om de skulle besvare henvendelser fra studenter. Innrapportering om studenter som har tatt slik direkte kontakt vil kunne bli regnet som fusk. Studentene oppfordres derimot til å besøke hjemmesiden til Oda samt butikker fra andre kjeder for å observere vareutvalg, priser, kvalitet, In-store markedsføring, kundeservice, utstilling (hvordan butikkene ser ut innvendig) etc, men det er ikke lov å intervjue de ansatte i forbindelse med oppgaven.

Om egne undersøkelser

Vi fraråder studentene å gjennomføre egne markedsundersøkelser i dette kurset. Dette kurset dekker ikke opplæring i GDPR, EUs regler for personvern som legger føringer på hvordan data skal håndteres og lagres. Kurset har heller ikke markedsundersøkelser som en del av pensum. Det er derfor ikke bare enkelt å lage en god og riktig spørreundersøkelse online. Videre er vår erfaring at slike egne undesøkelser ofte ikke styrker studentenes besvarelser (snarere motsatt), og de kan også ta fokuset bort fra bruk av pensum og den informasjonen som faktisk gis i caseteksten og vedlegg.

Relevant informasjon som skal brukes i analysene blir presentert i caseteksten. Videre oppfordres dere til å søke på internett og bruke bibliotekets databaser da dette gir viktig tilleggsinformasjon.

Prosjektoppgaven består av fire deler: Markedsplan, Faglig begrunnelse, Vedlegg og Egenevaluering. Markedsplanen og den faglige begrunnelsen teller 50 prosent hver av karakter på prosjektoppgaven (se også tabellen "Kriterier for vurdering av prosjektoppgaven" senere i dette dokumentet), og prosjektoppgaven teller 70 prosent av karakteren i kurset MRK 3414. Vedlegg og Egenevaluering teller ikke på karakteren.

Vedleggene kan imidlertid gi sensorene et fyldigere grunnlag for å vurdere arbeidet som er lagt ned i oppgaven. Dette gjelder spesielt vedlegg av utdrag fra øvelsesoppgavene.

1. Markedsplan

Markedsplanen skal utarbeides etter det oppsettet og de retningslinjer som er gitt i pensum til kurset. Markedsplanen er et dokument som skal leveres til bedriftens ledergruppe og styre som ett av flere dokumentgrunnlag for vedtak av strategi og budsjetter for virksomheten.

Markedsplanen skal være faktabasert, profesjonell, grundig og realistisk. I dette dokumentet er det viktig å vise evnen til å argumentere for løsningen, herunder evnen til å kommunisere begrensninger og forutsetninger løsningen bygger på.

Markedsplanen skal altså skrives slik man ville skrevet det som et ledelsesdokument.

Markedsplanen skal være på minimum 5 og maksimum 10 sider.

2. Faglig begrunnelse

Den faglige begrunnelsen skal beskrive hvordan man er kommet fram til de valg som er gjort i markedsplanen. I den faglige begrunnelsen skal man vise hvordan teorien brukes til å drøfte ulike forhold knyttet til bedriftens markedsføring. Man skal vise evnen til å anvende relevant teori, og man skal vise at man har en god forståelse av teorien. I dette ligger også evnen til ikke å trekke inn teori som ikke bidrar til å løse oppgaven.

Den faglige begrunnelsen skal bestå av to hoveddeler der den første delen omhandler analyse av bedriftens strategiske markedsutfordringer og den andre delen omhandler bedriftens strategiske markedsføringstiltak.

I markedsplanen har man definert de viktigste strategiske markedsutfordringer (eller markedsføringsproblemer) bedriften står overfor. Disse følger av analysen som dere gjør av markedet, konkurrentene, kundene og bedriftens markedsføring (herunder bedriftens lønnsomhet), opp imot de målene man har satt. I den faglige begrunnelsen skal man derfor redegjøre for hvilke analyser som er gjennomført og de viktigste funnene i disse, samt redegjøre for de målene som er gitt i markedsplanen. Dette skal lede frem til en argumentasjon for de utvalgte strategiske markedsutfordringene.

I markedsplanen har man kommet fram til de markedsstrategiske tiltak som skal løse bedriftens strategiske markedsutfordringer. I den faglige begrunnelsen skal man underbygge disse tiltakene med teori i kombinasjon med fakta og logiske resonnementer. Redegjørelsen skal utføres på et relativt overordnet nivå og uten at man blir for detaljert i hvordan tiltakene skal gjennomføres. Typisk vil en markedsplan inneholde 3-5 overordnede tiltak, og for hver av disse gis en faglig redegjørelse. Redegjørelsen skal også omfatte en begrunnelse for tiltakenes kostnader og for tiltakenes forventede effekt på inntekter og lønnsomhet (herunder hvilke forutsetninger resonnementet bygger på). Bruk av regneark, enten det som følger med boken eller et man utvikler selv, bidrar til å styrke argumentasjonen i den faglige begrunnelsen idet man viser hvordan tiltak vil bidra til å nå de overordnede målene.

I den faglige begrunnelsen er det viktig å redegjøre for kildene til teori, modeller og fakta man bruker. Kildehenvisninger skjer etter de retningslinjer som gjelder for BI studenter (biblioteket utgir informasjon om dette). Det er imidlertid ikke hensiktsmessig (eller ønskelig) å henvise til kilder i alt man skriver. Og det er dessverre ikke mulig å gi en klar og entydig retningslinje for når man skal henvise til en kilde, så her må man utvise et visst skjønn. Det er for eksempel ikke hensiktsmessig å angi en kilde når man argumenterer for at investeringer i distribusjon eller reduksjon i pris fører til økt salg. Det er imidlertid hensiktsmessig å angi en kilde dersom man for eksempel skulle argumentere for at en form for distribusjon har større effekt enn en annen, eller at man skulle argumentere for at priselastisiteten for en spesiell produktkategori er spesielt lav.

Den faglige begrunnelsen skal kunne leses uten å måtte lese vedleggene.

Markedsplanen og den faglige begrunnelsen skal være selvstendige dokumenter. Den faglige begrunnelsen skal være på minimum 10 og maksimum 15 sider.

3. Vedlegg

Vedlegg brukes til å dokumentere analyser man mener ikke hører hjemme i hovedteksten men som er sentrale for dokumentet. Vedlegg kan også brukes til å gjengi økonomisk analyse, og lignende. Øvelsesoppgavene legges inn som vedlegg idet dette vil gi sensor informasjon om hvilket arbeid som er lagt ned i prosjektoppgaven.

4. Egenevaluering

I egenevalueringen gis en kortfattet beskrivelse av læringsutbytte i forhold til kursets læringsmål. Egenevalueringen skal være maksimum 1 side.

Kriterier for vurdering av prosjektoppgaven:

Kriteria	**Beskrivelse**
Markedsplan: - Faktabasert	Finne fram til gode informasjonskilder. Utnyttelse av de fakta man har. Kunne ta rimelige forutsetninger. Sette sammen fakta fra ulike kilder. Trekke ut essensen og identifisere hovedproblemstillingene. Bredde i faktagrunnlag.
Markedsplan: - Profesjonell stil	Kortfattet. Presis. Korrekt. Lettlest. Engasjerende. Lettfattelige tabeller og figurer.
Markedsplan: - Grundig	Faktabasert. Gjennomtenkt. Gjennomdiskutert. Konkret og kontrollerbar.
Markedsplan: - Realistisk	Logisk. Forpliktende. Realistisk.
Faglig begrunnelse: - Prinsipiell drøftelse	Anvende teori, modeller og sjekklister fra pensum (inklusiv artikkelsamling) til drøftelse og analyse av utvalgte problemstillinger. Evne å trekke konklusjoner (essens) fra slike drøftelser, samt hva slike konklusjoner eventuelt betyr i praksis (i markedsplanen). Hvilke økonomiske konsekvenser (i budsjettet) de ulike tiltakene man planlegger har. Hvilke effekter (på merkekjennskap, omsetning, lønnsomhet osv.) disse tiltakene har.
Faglig begrunnelse: - Relevant teoribruk	Evne å anvende den teori (fra pensum inklusive artiklene) som er mest relevant i forhold til de problemstillinger bedriften stå overfor. Unngå referanser til teori man ikke bruker. Beherske bruk av BIs retningslinjer og standard for siteringer og referansebruk.
Faglig begrunnelse: - Forståelse av teori	Vise at man forstår hva teorien betyr. Riktig definisjon og bruk av begreper. Evner å se viktige nyanser i teorien og anvendelsen av denne. Skriver om teori på en enkel og lettfattelig måte som viser at man er trygg på hva teorien omhandler.
Faglig begrunnelse: - Dybde i teoribruk	Vise at man behersker pensum, og at man gjerne anvender kildene i pensum (studerer referansene og anvender disse).

Vurderingskriterier:

I overenstemmelse med de mer detaljerte kriteriene i tabellen ovenfor, som beskriver hva som kjennetegner en god markedsplan og faglig begrunnelse, er en A- eller B-besvarelse karakterisert av følgende:

Markedsplan:

Faktabasert, profesjonell, grundig og realistisk

Faglig begrunnelse:

Utstrakt bruk av relevant teori og fakta i sin analyse og tiltak. For eksempel ta hensyn til at Oda har ulike kundesegmenter som antakelig fordrer ulike markedsstrategier.

Vise til tiltak og økonomiske konsekvenser av disse (tiltakene) basert på analysen. Og, i denne sammenheng bruke regneark/økonomisk oppstilling (basert på malen vi har benyttet i øvelsesoppgavene eller egenprodusert).

Diskutere/vise til sammenhengen mellom tiltakene og ulike effekter. For eksempel, sammenhengen mellom kommunikasjon, merkestyrke og omsetning.

En tydelig rød tråd i oppgaven som binder de ulike delene sammen. Faglig begrunnelse skal utdype og forklare markedsplanen (gi merverdi). Analysen dere gjennomfører skal lede oss frem til de strategiske markedsutfordringene som i sin tur setter premissene for hvilke mål dere setter for Oda. Valg av markedsstrategi (og tiltak) må svare til målene som er satt og det er forventet at dere synliggjør effektene av tiltakene. Kritisk tankegang og gode selvstendige resonnementer (basert på fakta og teori) er nøkkelen til en god karakter.

Karakterer:

A = Fremragende
Karakterbeskrivelse: Fremragende prestasjon som klart utmerker seg. Kandidaten viser svært god vurderingsevne og stor grad av selvstendighet.

B = Meget god
Karakterbeskrivelse: Meget god prestasjon. Kandidaten viser meget god vurderingsevne og selvstendighet.

C = God
Karakterbeskrivelse: Jevnt god prestasjon som er tilfredsstillende på de fleste områder. Kandidaten viser god vurderingsevne og selvstendighet på de fleste områder.

D = Nokså god
Karakterbeskrivelse: En akseptabel prestasjon med noen vesentlige mangler. Kandidaten viser en viss grad av vurderingsevne og selvstendighet.

E = Tilstrekkelig
Karakterbeskrivelse: Prestasjon som tilfredsstiller minimumskravene, men heller ikke mer. Kandidaten viser liten vurderingsevne og selvstendighet.

F = Ikke bestått
Karakterbeskrivelse: Prestasjon som ikke tilfredsstiller de faglige minimumskravene. Kandidaten viser både manglende vurderingsevne og selvstendighet.

Vedlegg 1 – casetekst.

NB! Av konkurransehensyn og vern om bedriftsinterne forhold er flere av tallene som presenteres i oppgaven endret/manipulert. Dere forutsetter likevel at tallene i oppgaven er reelle og at dere brukes disse i analysearbeidet. NB! Husk at oppgaven kun skal omhandle det norske markedet, ikke internasjonalt.

Oda

Oda er Norges største matbutikk på nett og har store vekstambisjoner både i Norge og i utlandet. I det norske markedet leverer vi til store deler av Østlandet (rundt 40% av Norges befolkning) og både til privatkunder og bedriftskunder. Anslagsvis 1 700 000 potensielle kunder bor i den delen av Norge hvor Oda leverer varer i dag. Under pandemien har bedriftsmarkedet blitt en betydelig mindre andel av omsetningen vår. Vi kommer til å fortsette å levere til små og mellomstore bedrifter i Norge, men vårt store satsningsområde er privatkunder og med hovedvekt på vår kjernemålgruppe barnefamilier. Vi er spente på å høre hvordan studentene på BI tenker at Oda skal vokse videre og ta større markedsandeler i Norge nå når vi er på vei inn i massemarkedet på adopsjonskurven.[1]

Oda (tidligere Kolonial.no) ble grunnlagt i 2013 av 10 venner fra Oslo med teknologi- og logistikkbakgrunn. Oda omsatte for nærmere 2 milliarder norske kroner i 2020 og er et av Europas raskest voksende selskaper. Oda har som mål å omsette for 10 milliarder kroner innen 2026. De har ikke satt noen dato for når de skal bli lønnsomme da deres fokus de neste årene vil være å vokse selskapet så raskt som mulig internasjonalt.

Oda har et mål om å gjøre hverdagen enklere for folk. Med en revolusjonerende teknologi og verdikjede er de allerede godt på vei. Oda har et utvalg som et supermarked, med priser som i lavprisbutikkene og dette er en posisjonering de ønsker at merket skal bli assosiert med. De har sitt eget bakeri og samarbeider med store og små merkevarer. Du finner også hundrevis av skreddersydde oppskrifter og middagstips på nettsidene til Oda.

Oda er først og fremst en aktør på dagligvaremarkedet i 2021, men kan også betraktes som en handelsplattform som selger varer fra ulike bransjer. I tillegg til dagligvarer har Oda samarbeidsavtaler med bedrifter fra andre bransjer om å selge deres varer via plattformen til Oda.

Verdier:[2]

Vi bryr oss om mennesker.

Vi jobber for å gjøre livet enklere for kundene våre. Vi er vennlige og tilgjengelige, men tar ikke opp unødvendig plass.

---

[1] Åpningskommentar fra ledelsen i Oda

[2] Verdiene til Oda er hentet fra deres nettsider

Sikter mot stjernene

Vi er et ambisiøst selskap som jobber for å levere hverdagsmagi. Vi har et stort fokus på utvikling og forbedring, fordi vi vet at bedre alltid er mulig.

Bygger noe som skal vare

God service og høy effektivitet betyr at man ikke kan ta snarveier. Vi ønsker langsiktige forhold til kundene våre og jobber for å være en pålitelig og trygg leverandør. I tillegg tar vi klima og miljø på alvor og vil bygge noe som er bærekraftig i samfunnet. Les om målene våre her: https://sustainability.oda.com/

Markedet:

Dagligvarebransjen i Norge omsatte for over 200 milliarder kroner i 2020. Bransjen består av mange kjeder (merker), men de fleste kjedene er konsentrert i tre grupperinger; NorgesGruppen ASA, Coop Norge SA og REMA 1000 Norge AS. Disse tre grupperingene innehar rundt 95% markedsandel i dagligvarebransjen i Norge. Kritiske røster påpeker at en slik maktkonsentrasjon har gjort det vanskelig for nye kjeder å etablere seg i Norge. På oppdrag fra Nærings- og fiskeridepartementet utarbeidet Oslo Economics i 2017 en rapport om «Etableringshindringer i dagligvaresektoren» som beskriver dette i detalj[3]. Dere finner rapporten på bibliotekets hjemmeside for dette kurset, men de viktigste konklusjonene fra studien var som følger:

• Det norske dagligvaremarkedet kjennetegnes ved høy grad av vertikal integrasjon, konsentrert leverandørindustri, lavt vareutvalg og høy butikktetthet.

• Stordriftsfordeler i innkjøp gjør at en nyetablerer må oppnå høye volumer på kort tid for at det skal være lønnsomt å etablere seg.

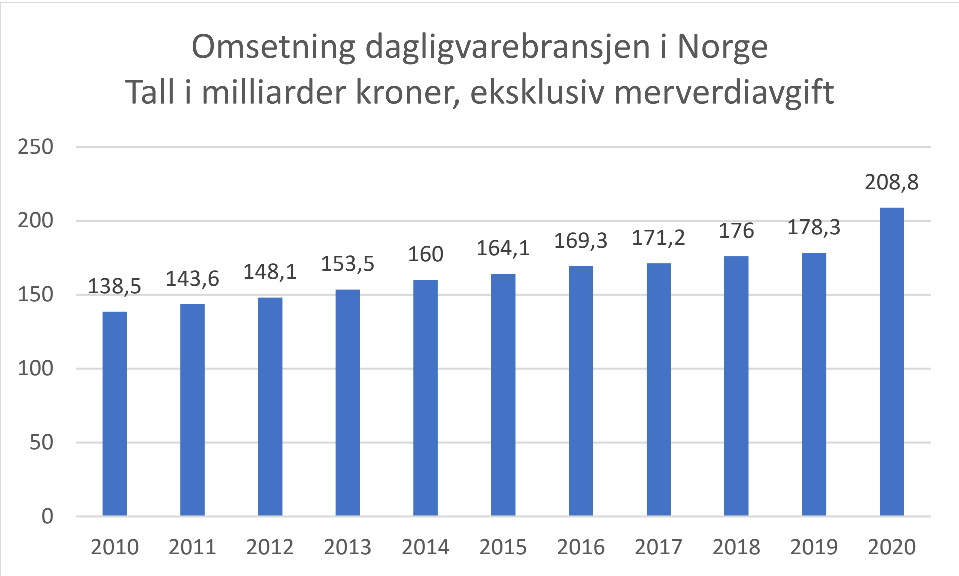
• Mulighetene for å vokse raskt begrenses av høy butikktetthet og sterk konkurranse om de mest attraktive lokalene.

• Økt konkurranse på leverandørleddet vil sannsynligvis redusere stordriftsfordeler i innkjøp og dermed gjøre det mer attraktivt å etablere grossist- og detaljistvirksomhet i Norge.

[3] «Etableringshindringer i dagligvaresektoren», Oslo Economics 2017

• Tiltak som åpner for økt konkurranse fra utlandet på leverandørleddet og lempning av konkurransebegrensende regulering i landsbrukssekstoren er trolig tiltak som vil ha størst effekt på sannsynligheten for etablering av grossist- og detaljistvirksomhet.

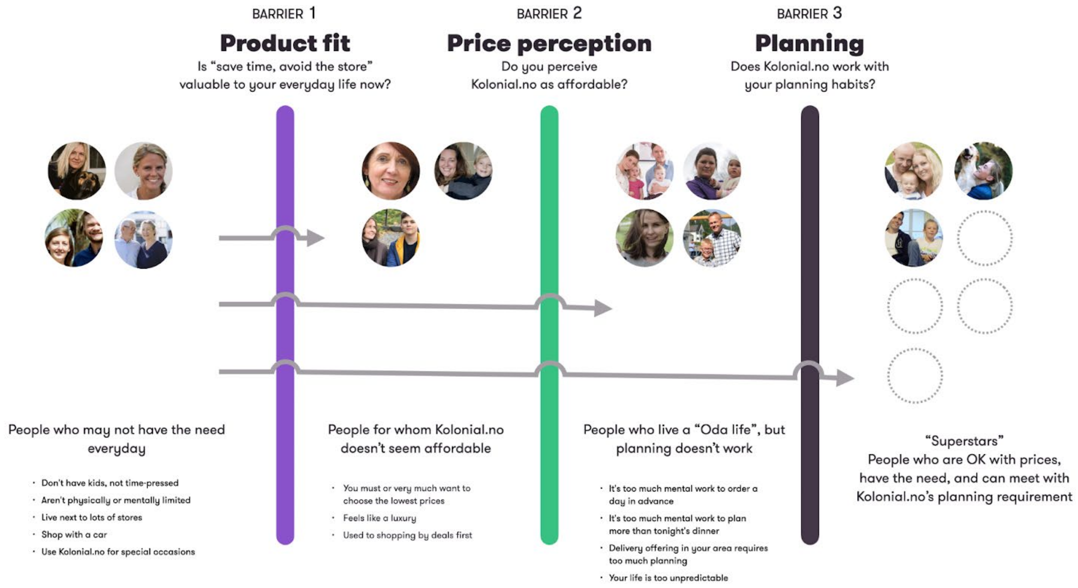
• Til dags dato synes aktører med alternative distribusjonsformer, særlig fullsortiments dagligvarer på nett, å ha størst potensial til å utøve et konkurransepress av betydning på den tradisjonelle dagligvarehandelen. Fremtidig utvikling i denne delen av dagligvaresektoren er imidlertid usikker.



Figur 1, omsetning dagligvarebransjen i Norge. Kilde: NielsenIQ

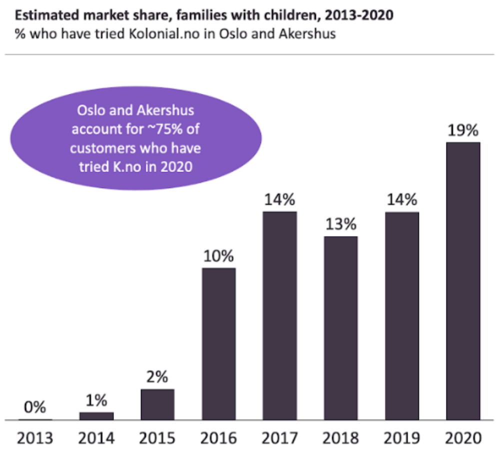
Dagligvarehandelen i Norge foregår hovedsakelig i fysiske butikker (anslagsvis rundt 98% av omsetningen i bransjen), men netthandelen er i vekst og opplevde en kraftig omsetningsvekst under pandemien. Oda doblet nesten omsetningen fra 2019 til 2020.

Dagligvarekunder i Norge har godt innarbeidede rutiner/vaner med å handle i fysiske butikker. For å øke andelen som handler dagligvarer på nett blir det derfor viktig for Oda å forsøke å få kundene til å endre sine vaner. I en undersøkelse rettet mot eksisterende kunder identifiserte Oda tre hovedbarrierer som hindrer kundene i å handle matvarer mer regelmessig på nett. Figur 2 oppsummerer analysen.

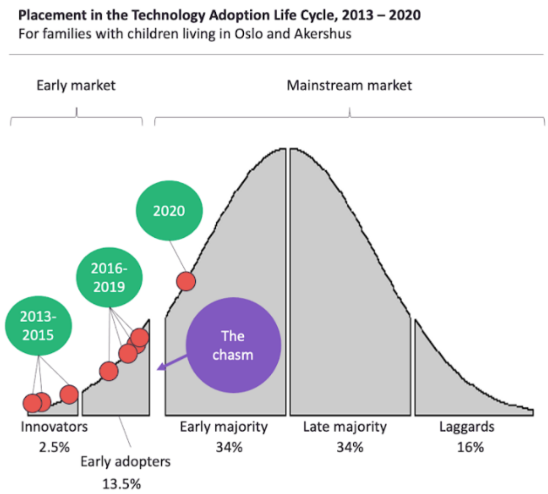


Figur 2, barrierer for å handle matvarer mer regelmessig på nett

Dagligvarehandel på nett er i en tidlig fase i sin livssyklus og har foreløpig en beskjeden markedsandel (1-2%) i dagligvarebransjen. I forhold til Oda sine beregninger har dagligvarehandel på nett, i Oslo og Akershus, nettopp kommet over i kategorien «tidlig majoritet» for kundesegmentet «husholdninger med barn», (figur 3 og 4). Det er naturlig å anta at andre geografisk områder i Norge, samt andre kundesegmenter, ennå ikke har kommet til «tidlig majoritet» i livssyklusen for dagligvarer på nett. I 2020 hadde Oda ca 75% markedsandel av dagligvarehandel på nett, mens meny.no hadde ca 20% og andre aktører ca 5%.



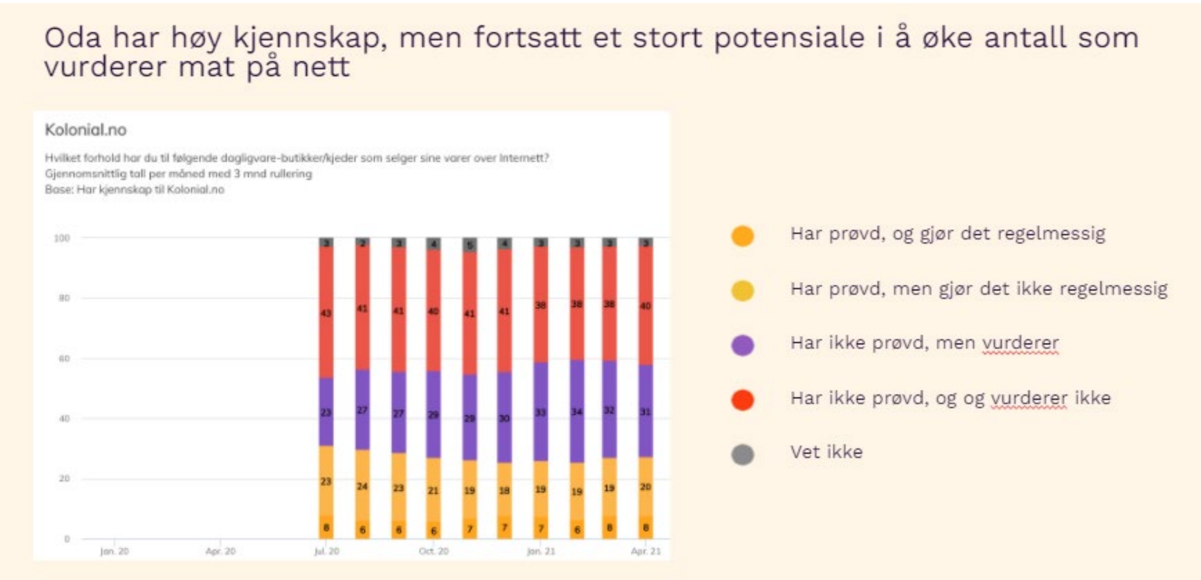
Figur 3 Prøvd Oda i Oslo og Akershus



Figur 4 Adopsjonskategorier

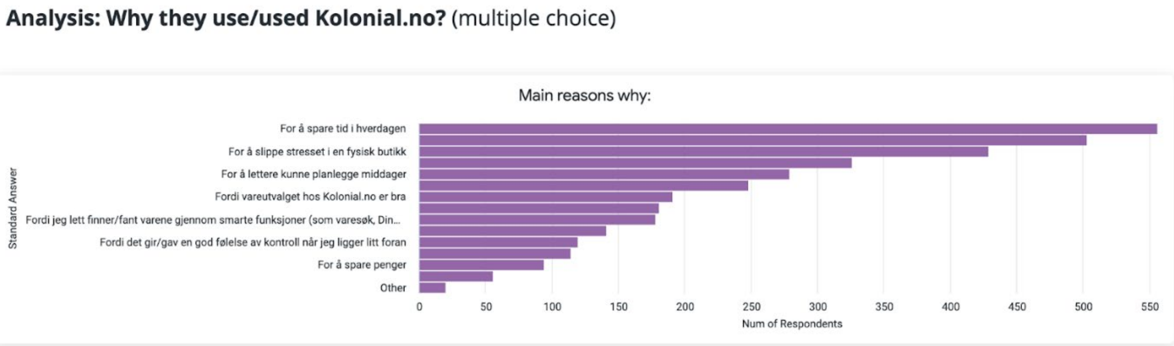
Oda har ambisjoner om ytterligere vekst og et nødvendig bidrag til å nå vekstmålene er å rekruttere nye kunder. Figur 4 viser at dagligvarehandel på nett akkurat har kommet inn i den tredje adopsjonskategorien så potensialet for ytterligere vekst er absolutt til stede. Ifølge adopsjonsprosessen så er første kriterium for å tiltrekke seg nye kunder at kundene faktisk har kjennskap til bedriften. Kjennskapen i markedet kan økes gjennom reklame og vareprat fra personer som allerede kjenner til bedriften. Oda gjennomførte en undersøkelse blant helt nye kunder hvor de spurte disse om hvor de hadde hørt om Oda (de kunne svare på flere kategorier). 57% av kundene svarte at de fikk vite om Oda gjennom en venn, 22% hadde sett reklame på TV eller utendørsreklame, 19% hadde hørt om Oda fra en kollega, 15% hadde hørt om Oda via facebook, 3% gjennom Instagram og 7% svarte «annet». Tallene viser viktigheten av word of mouth (vareprat) i arbeidet med å rekruttere nye kunder.

Det er flere steg kundene må gjennom i adopsjonsprosessen før de blir aktive regelmessige kunder. Figur 5 viser en oversikt over hvordan forbrukere som kjenner til Oda fordeler seg i de etterfølgende stegene i adopsjonsprosessen



Figur 5: Vurdering, prøving og lojalitet

Det kan være mange årsaker til å handle dagligvarer på nett. I en undersøkelse Oda gjennomførte blant sine egne kunder oppgav flest respondenter at det var «for å spare tid i hverdagen» etterfulgt av «å slippe stresset i en fysisk butikk» (se figur 6).



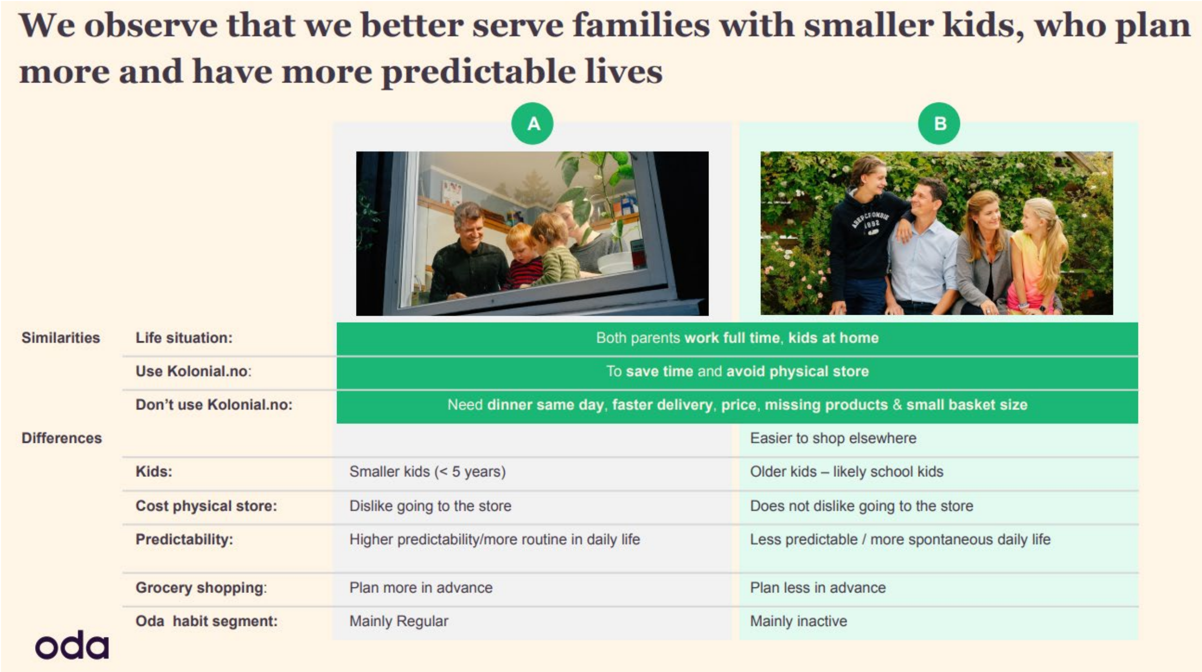
Figur 6: Årsaker til å ha brukt Oda (Kolonial.no). Totalt 961 personer deltok i undersøkelsen

Segmenter

I denne oppgaven har vi definert tre makrosegmenter for Oda. Disse kan igjen deles inn i mikrosegmenter. Segmentinndelingen er gjort med tanke på ulik kjøpsadferd og deles inn i 1) Forbrukermarkedet (B2C) husholdninger u/barn, 2) B2C husholdninger m/barn og 3) B2B (bedriftsmarkedskunder).

Forbrukermarkedet står for ca 90% av omsetningen til Oda og den viktigste målgruppen er husholdninger med hjemmeboende barn. Dette segmentet utgjør ca 35% av kundene som har prøvd Oda, men de står for over 50% av omsetningen (2020) på forbrukermarkedet. Tabell 5 viser økonomien i de tre segmentene.

Hovedmålgruppen «husstander med barn» kan deles inn i mikrosegmenter, husholdninger små barn (under 5 år), og husholdninger med større barn. Oda har erfart at det er forskjeller basert på hvor gamle barna er. Figur 7 beskriver forskjeller og likheter mellom to mikrosegmenter med tanke på kjøpsadferd.



Figur 7, Beskrivelse av mikrosegmentene husholdninger med små/store barn

B2B

Bedriftskunder har de siste 3 årene representert 10-20% av Oda´s totale omsetning. Bransjefordelingen av kundene er ca 30% offentlig sektor (kommunale kontor, helsetjenester, barnehager og skoler), 15% kunder innen bygg og anlegg, 5% barnehager og de resterende kundene er SMB-kontorer.

Konkurrentene i dette segmentet er fysisk dagligvare (Kiwi, Coop, Rema 1000 etc), netthandel (Meny, Coop), spesialister (Lunsj.no, Helt Opplagt, Toolbox, Cater, med fler) og kantiner, kafeer, foodora osv.

Oda har primært fokus på bedrifter som har like god product/market fit som privatkunder. Den primære målgruppen er selskaper som har mellom 10-100 ansatte på en lokasjon. Oda har pr i dag sterk markedspenetrasjon i offentlig sektor, barnehager og bygg og anlegg, men har stort potensiale i "vanlige" kontorlandskap.

Pandemien har ført til store endringer i dette segmentet og omsetning har falt med ca 60% under Covid-19. Et viktig bidrag for å vinne tilbake den tapte omsetningen blir reaktivering av kunder. Faste lunsjkunder er de mest lukrative kundene til Oda, og de ønsker å ta en posisjon som markedsledende lunsjleverandør.

Markedsaktiviteter i B2B-segmentet fordeles likt 50/50 mellom telemarketing (nysalg og reaktivering/winback) og aktiviteter på digitale flater (Google ads, LinkedIn og annonser/reklame på Facebook, E24, print)

Produktkategorier

Oda har inndelt produktene i tre hovedkategorier; 1) ferskvarer (frukt og grønt, kjøtt, fisk, fugl, brød m.m.), 2) tørr-/frysevarer og 3) non-food. De to første kategoriene omhandler tradisjonell dagligvarehandel, mens non-food inneholder produkter fra andre bransjer. I den tredje kategorien (non-food) har Oda samarbeidsavtale med Clas Ohlson, Sprell, Barnas Hus og Mester Grønn som innebærer at Oda selger deres produkter via sin handelsplattform. Økonomien i de tre kategoriene får dere oversikt over i tabell 5.

Oda tilbyr er bredt vareutvalg og har blant annet en avtale med REMA distribusjon som foretrukket leverandør. Avtalen innebærer at Oda også får tilgang til varer under varemerker REMA 1000 eier eller har eksklusiv bruksrett til. Oversikt over vareutvalget og priser finner dere på Oda sine nettsider.

Markedsføring

Oda markedsfører seg i forskjellige kanaler. Ca 30% av markedsføringsbudsjettet blir brukt i offline channels (TV og utendørsreklame) og 70% i online channels.

I online channels fordeles budsjettet slik:

30% Facebook/Instagram

40% Søkeoptimalisering (Google Ads)

10% Display banner Premium sites (VG, Nettavisen, DN, etc.)

20% Online video Premium sites YouTube

Markedsundersøkelser:

Oda måler jevnlig markedets kjennskap, vurdering og preferanse av dagligvare-butikker/kjeder som selger sine varer over internett. Resultatene fra undersøkelsen i mars 2021 ser du nedenfor. Undersøkelsen er utført før de byttet navn til Oda og gjelder derfor fra tiden de het Kolonial.no. Undersøkelsen er gjennomført i de geografiske områdene Oda leverer mat/produkter.

Spørsmål:

Uhjulpen kjennskap:	Hvilke dagligvare-butikker/kjeder som selger sine varer over internett, kjenner du til eller har du hørt om?
Hjulpen kjennskap:	Hvilke av følgende dagligvare-butikker/kjeder som selger sine varer over Internett kjenner du til eller har du hørt om?
Vurdering:	Hvilke kjeder/butikker ville du vurdert å benytte dersom du skulle kjøpe dagligvarer fra Internett?
Preferanse:	Dersom du skulle kjøpe dagligvarer fra internett, hvilken kjede/butikk ville du mest sannsynlig valgt?

	Kolonial.no (nå Oda)	Meny.no	Morgenlevering	Spar.no
Uhjulpen kjennskap	58 %	71 %	1 %	4 %
Hjulpen kjennskap	90 %	82 %	55 %	28 %
Vurdering	61 %	47 %	32 %	9 %
Preferanse	43 %	26 %	9 %	2 %

Tabell 1

Når Kolonial.no byttet navn til Oda medførte dette selvsagt endringer i tallene. Et nytt merkenavn vil ta tid å innarbeide. Oda utfører disse undersøkelsene månedlig. I spørsmålene for hjulpen kjennskap, vurdering og preferanse er Oda beskrevet slik: Oda (tidligere Kolonial.no). Tabellen nedenfor viser utviklingen før/etter navnebytte:

	Kolonial.no (mars)	Oda (april)	Oda (mai)	Oda (juni)	Oda juli)
Uhjulpen kjennskap	58 %	2 %	9 %	15 %	23 %
Hjulpen kjennskap	90 %	66 %	70 %	72 %	76 %
Vurdering	61 %	32 %	41 %	47 %	49 %
Preferanse	43 %	21 %	27 %	31 %	33 %

Tabell 2

Norsk Kundebarometer (NKB) og Norsk Bærekraftbarometer (NBB) gjennomfører årlige undersøkelser av norske bedrifter på forbrukermarkedet. Bedriftene blir målt på kundetilfredshet, lojalitet (gjenkjøpssannsynlighet) og bærekraft. Nærmere informasjon om undersøkelsene finner dere på www.bi.no/nkb og www.bi.no/nbb. Tabellen nedenfor viser resultatene innenfor dagligvarebransjen 2021. Husk at Oda også selger produkter fra andre bransjer.

Kilde ->	bi.no/nkb	bi.no/nkb	bi.no/nbb
Selskap	Tilfredshet	Lojalitet	Bærekraftscore
Oda (kolonial.no)	78	86,2	63,6
Meny	77,7	87,3	67,7
Obs	74,7	89,9	70,2
Spar	74,4	90,6	69
REMA 1000	73,6	90,9	63,8
Kiwi	73,1	90	65,6
Extra	72,5	88,9	68,5
Coop Mega	71,2	86,4	68
Coop Prix	63,4	83,7	65,2
Joker	62,7	84,7	62,5
Bunnpris	60,7	82,2	55,6

Tabell 3

Økonomi Oda:

RESULTATREGNSKAP i hele 1000	**2020**	**2019**	**2018**	**2017**	**2016**
Valutakode	NOK	NOK	NOK	NOK	NOK
Sum salgsinntekter	1 977 731	1 095 058	919 334	801 394	424 021
Annen driftsinntekt	643	-	-	0	0
Sum driftsinntekter	**1 978 374**	**1 095 058**	**919 334**	**801 394**	**424 021**
Varekostnad	1 361 922	749 247	644 487	599 786	337 764
Lønnskostnader	231 287	177 590	174 819	191 336	93 691
Ordinære avskrivninger	26 776	22 968	17 896	12 490	6 187
Andre driftskostnader	464 859	302 032	255 175	251 315	103 793
Driftsresultat	**-106 470**	**-156 779**	**-173 042**	**-253 531**	**-117 415**

Note: Regnskapstall for Oda (Kolonial.no) 2016 – 2020. Tall i tusen kroner. Hentet fra PROFF

Tabell 4: Regnskapstall Oda

Tabell 5

Tabell 5: Inntekter og kostnader Oda 2020, fordelt på segment og produktkategori

Kommentar til tallene: Dere vil observere at tallene i tabell 5 avviker noe fra resultatregnskapet i tabell 4 hentet fra Proff. Årsaken er at tallene i tabell 5 ikke inneholder alle tall for selskapet, men gjelder de tre segmentene og de tre produktkategoriene vi har valgt å fokusere på i dette caset.

Når dere skal vurdere historisk utvikling for Oda og markedsandeler kan dere benytte tabell 4, mens for analyser dere skal benytte regnearket til, benytter dere tall fra tabell 5 (det er forventet at studentene benytter regnearket som medfølger boken til å vurdere lønnsomhet i segmenter/kategorier, fremtidig utvikling, effekter og kostnader av tiltak m.m.).

År -1		Tørr/frys			Ferskvarer			Non food		
	Totalt	Husholdning u/barn	Husholdning m/barn	B2B	Husholdning u/barn	Husholdning m/barn	B2B	Husholdning u/barn	Husholdning m/barn	B2B
Pris pr vare		32,90	29,52	33,87	30,54	29,86	30,57	34,81	35,22	32,62
Antall enheter solgt		9 665 497,12	10 958 856,95	2 690 814,00	15 135 221,72	15 480 244,95	3 550 146,10	1 863 609,51	1 857 898,50	271 291,00
Salgsinntekter	**1 904 751 566**	317 994 855	323 474 284	91 130 217	462 229 671	462 229 671	108 540 301	64 872 739	65 430 109	8 849 718
Faste kostnader	**213 958 888**	38 446 933,60	38 446 933,60	9 947 790,08	50 177 410,49	50 177 410,49	11 769 190,07	7 016 815,64	7 016 815,64	959 588,40
Variabel enhetskostnad (pr. handlekurv/solgt)		29,08	26,48	29,29	25,90	25,65	26,26	29,90	29,99	28,02
Variable enhetskostnader	**1 651 332 585**	281 118 356	290 182 409	78 801 104	391 930 823	397 026 105	93 229 266	55 721 587	55 721 587	7 601 349
Driftskostnader	**1 865 291 473**	319 565 289	328 629 342	88 748 894	442 108 233	447 203 515	104 998 456	62 738 403	62 738 403	8 560 937
Driftsresulat før salg og markedsføring	**39 460 093**	-1 570 434	-5 155 058	2 381 323	20 121 438	15 026 156	3 541 845	2 134 336	2 691 706	288 781
Faste markedsføringskostnader	**20 750 685**	3 728 754,70	3 728 754,70	964 780,95	4 866 428,54	4 866 428,54	1 141 428,42	680 522,00	680 522,00	93 065,15
Variable markedsføringskostnader pr.handlekurv solgt		0,79	0,72	0,80	0,71	0,70	0,71	0,81	0,82	0,76
Variable markedsføringskostnader	**44 986 607**	7 680 233	7 900 701	2 145 492	10 670 972	10 809 700	2 538 323	1 517 113	1 517 113	206 959
Rekalmeinvesteringer	**30 763 754**	5 528 033,91	5 528 033,91	1 430 327,88	7 214 682,71	7 214 682,71	1 692 215,11	1 008 902,17	1 008 902,17	137 972,96
Sum markedsføringskostnader	**96 501 045**	**16 937 022**	**17 157 490**	**4 540 601**	**22 752 083**	**22 890 811**	**5 371 966**	**3 206 538**	**3 206 538**	**437 998**
Driftsresultat	**-57 040 953**	**-18 507 456**	**-22 312 548**	**-2 159 277**	**-2 630 645**	**-7 864 655**	**-1 830 121**	**-1 072 201**	**-514 832**	**-149 217**
	-2,99 %	-5,82 %	-6,90 %	-2,37 %	-0,57 %	-1,70 %	-1,69 %	-1,65 %	-0,79 %	-1,69 %

Vedlegg til oppgaven

Biblioteket hjemmeside for Prosjektoppgaven

På disse sidene finner dere diverse kilder og statistikk med relevans for denne prosjektoppgaven. Blant annet vil dere finne:

- Virke Handelsrapporten 2020-2021
- Etableringshindringer i dagligvaresektoren
- Netthandelsbarometeret juni 2021
- Dagligvarefasiten 2021
- Aktørbildet i dagligvarekjeden, notat Menon 30 oktober 2020

Link til studentportalens side med formelle krav, oppgavemaler, innlevering:

https://portal.bi.no/eksamen-og-oppgave/oppgaveskriving/formelle-krav/

Nettartikler:

https://dagligvarehandelen.no/nyheter/2021/avslutter-sak-mot-orkla-mondelez-og-norgesgruppen

Nets_Norsk e-handel 2020.pdf

https://www.menon.no/wp-content/uploads/2018-77-Fjerning-av-350-kronersgrensen-for-import-til-Norge.pdf

https://no.ehandel.com/meny-no-vokser-men-vi-tjener-ikke-penger-pa-nett

https://www.ssb.no/varehandel-og-tjenesteyting/artikler-og-publikasjoner/detaljhandelen-okte-med-11-prosent-i-2020

MRK 34143 Markedsføringsledelse (skriftlig eksamen)

Termin: Høst 2021
Lengde på besvarelse: veiledende 4 sider
Maks antall vedleggsfiler til besvarelsen: 0
Oppgaven besvares individuelt

Formelle krav:
Du trenger ikke å henvise til kilde eller lage litteraturliste, men benytt gjerne skrifttype, linjeavstand og marger som i BIs mal for oppgaveskriving

> Marger: 5 cm venstremarg, 2 cm høyre-, topp- og bunnmarg
> Linjeavstand: 1,5
> Skrifttype- og størrelse: Times New Roman 12 pkt (eller tilsvarende, f.eks. Calibri 11 pkt)

Du har 2 timer på å løse oppgavene. Det forventes at du bruker pensum til å løse oppgavene. Det er ikke tillat å samarbeide med andre. Oppgaven skal besvares individuelt. Hver deloppgave teller 25% (totalt 100%).

Oppgave 1

a) Forklar begrepene makrosegmentering og mikrosegmentering. Bruk eksempler.
b) Hva er en merkevare, og hvilke roller har den? Bruk eksempler.

Oppgave 2:
Du og en venn skal starte opp en pub i sentrum av en stor norsk by til våren.
Denne puben skal være en type tradisjonell irsk pub, som er åpen alle dager fra kl 12-01, med livemusikk fredag og lørdag. Lokalet dere har fått ligger sentralt med tanke på uteliv.
Pubens viktigste inntektskilde er det som selges fra tappekranene. Dere planlegger å ha 10 tappekraner i puben. Dere skal nå ha et møte om prising. På generelt grunnlag:

a) Hvilke forhold vil dere diskutere som kan ha betydning for hvordan dere setter prisene på ølet dere skal selge fra tappekranene? (Argumenter i forhold til pensum, og ta egne forutsetninger der det trengs).
b) Hvilke andre attributter enn pris kan være viktig for å skape nytte for kundene til denne type virksomhet? (Tenk gjerne på hva som kan påvirke kundetilfredsheten)

STR 36053 Strategi

Semester:	Høst 2021
Besvares:	I grupper 1-3
Besvarelsens omfang:	Maks 6000 ord, ekskl. vedlegg

Prosjektoppgaven skal bestå av en strategisk analyse og en anbefaling av strategi for en selvvalgt bedrift eller frivillig organisasjon.

Oppgaven legger vekt på at man kan beskrive og analysere bedriften godt - og at man kan tilføre kunnskap som kan lede til beslutninger om virksomhetens fremtid. Analyserapportenskal lede frem til en presentasjon til en tenkt ledergruppe eller styre. Konklusjonene og drøftingene skal gjenspeile at man evner å bruke relevant teori og analyseverktøy fra strategifaget.

Den anbefalte strategien bør være forankret i strategihjulet hvor det kommer frem hvilke leveranser man skal ha på hvilke markeder, hvilke aktiviteter og ressurser som understøtter dette, og hva som kreves for å gå fra nåværende situasjon til fremtidig ønsket situasjon. Det legges betydelig vekt på at det strategiske forslaget er realistisk og gjennomførbart, at man vurderer risikoen ved den valgte strategien.

Prosjektoppgaven består av:

1. En analyserapport på maks 6000 ord med Executive Summary på maksimalt 1-en A4-side (300 ord i tillegg). En utførlig forklaring om sitering og referanseteknikk, finner du på bibliotekets hjemmeside: https://portal.bi.no/eksamen-og-oppgave/oppgaveskriving/vise-til-kilder/. Krav til layout og lenke til veiledende mal finner du på www.bi.no/veiledendemal.
2. Powerpoint presentasjon for styret eller ledelsen i virksomheten (maks 10 slides). Powerpoint-slides kan settes opp med to slides pr side – uten kommentarer.

Følgende er viktig:

- Velg et foretak som faktisk kan gjøre strategiske valg. Det kan oftest ikke enkeltvirksomheter innen kjedeforetak som helse, sport, dagligvarer etc. Da er det bedre å skrive om hele kjeden. Se gjerne video for råd om valg av bedrift.
- Det er viktig at gruppen velger foretak innen 10. september.
- Det er bedre å skrive flere sammen enn å skrive alene. Vi tillater ikke grupper påflere enn tre personer.
- Det er en stor fordel om du kjenner noen som jobber i virksomheten, eller har god tilgang til ledende personer for intervjuer. Det hjelper på forståelsen og analysene, og mulighetene for å finne realistiske anbefalinger som kan gjennomføres i praksis.