*Classification of skin diseases Using Machine learning*

Frank Kvarme Faanes

Gabriel Dähn Wiland

Table of Contents

[1. Introduction 1](#_Toc99955248)

[2. Background 1](#_Toc99955249)

[3. Problem statement 1](#_Toc99955250)

[4. Project objectives 2](#_Toc99955251)

[5. Data warehouse details 2](#_Toc99955252)

[6. Interesting insights 2](#_Toc99955253)

[7. Data pre-processing 2](#_Toc99955254)

[8. Measuring performance 2](#_Toc99955255)

[9. Algorithms applied 3](#_Toc99955256)

[9.1 Algorithm 1 3](#_Toc99955257)

[9.2 Algorithm 2 3](#_Toc99955258)

[9.3 Algorithm 3 3](#_Toc99955259)

[9.4 Comparison 3](#_Toc99955260)

[10. Discussion 3](#_Toc99955261)

[11. Conclusion 3](#_Toc99955262)

[References 4](#_Toc99955263)

1. **Introduction**

The introduction to any report leads the reader into your topic and should grab their attention. It should consist of a paragraph or two from which the user may gain insight as to what the report is about, how it will derive its findings and why the reader should be interested.

* Hvorfor programmet kan være et verktøy, i forhold til dagens behandlingstider

1. **Background**

A description of the overall domain in which your project/problem is founded. For instance, if your project is based in the domain of online marketing, then you need to provide background information related to online marketing. This is important since the reader needs to understand the overall domain of your study to understand the significance of the problem and the contribution of your solution.

* Bakgrunnsinfo om hvorfor dette kan være nyttig (dagens metoder)
* Ressurskrevende å diagnosere
* Potensiell lang behandling på grunn av feildiagnoser
* Hvorfor er programmet nyttig, beskriv for leser
* Stor vekst i AI og viktig å finne løsninger i alle bransjer
* Litt statistikk om hvor vanlig disse sykdommene er?
* Vanskelighet å gi diagnose på grunn av stor likhet på ulike sykdommer (vise til eksempel)

The background section could contain several sub-sections if there are sub-topics that the reader should understand, to grasp the main topic. A good rule of thumb to the question of “how much should I write”: If you were a person who knew absolutely nothing of a topic, but you were expected to do work based on knowledge of this topic, how much would someone need to tell you about the topic. In this scenario, you are the person who needs to convey the information about the topic, so that someone else has a working knowledge of it.

* Forklar de ulike klassene som blir klassifisert (sopp, virus infeksjon, bakteriell)
* Potensiell behandling, hvor raskt må man behandle sykdommen
* Tidlig oppdagelse gir bedre prognose?

In this section, and any other, you need to provide citations whenever you use work originally created by others. These works include paraphrased text, directly quoted text, images (whether directly used or modified), tables, code and equations. In the case of text, add the citation as close to the cited text as possible. In the case of images, tables, code and equations, the citation should be added to the caption. Remember that every citation requires an associated entry in the report’s reference list. At Noroff, we make use of the Harvard referencing and citation style. If you are unsure of how to use this citation style, visit the following link: <https://www.mendeley.com/guides/harvard-citation-guide/>

When including images, tables, code and/or equations, remember to provide these elements with numbered captions. The reason for providing numbered captions is so that you can reference the item in text, so that the attention of the reader may be drawn to it. As a general rule of thumb, if a graphical element (whether image, table, code or equation) is not referenced in the text, then it should not be included in your report. For instance, if you have a Figure 1 somewhere in your report (preferably close-by to where it is referenced in text), then you need to have a statement somewhere in you your body of text that states something similar to: “Figure 1 demonstrates….” or “…this process is graphically represented in Figure 1.”.

1. **Problem statement**

In this section, it is assumed that the reader already understands the domain of your problem. Now it’s time to focus the reader’s attention on the specific aspect of the overall problem domain that you want to address. A problem statement is normally a succinct (brief and to the point) statement regarding the problem. Remember, it should just state the problem and not how to solve it; the “solving” part is what your study/report is about. The following is an example problem statement:

“It is difficult to predict the croaking duration of tree frogs.”

* “Long processing time and risk of wrong diagnosis because of human error”

1. **Project objectives**

* Create efficient models that can classify and distinguish diseases.
* Research different methods of extracting useful information from images.

The main objective of your study serves to state what the plan is for addressing the problem statement. It may consist of multiple sub-objectives (if you choose). Sub-objectives may be seen as sub-goals to be addressed to address the main study objective. An objective is written in terms of a single action to be completed. It should be concise and not descriptive of the task associated with the objective. The following is an example objective:

“Devise a predictive model to predict the croaking duration of tree frogs.”

1. **Data warehouse details**

* Fra Kaggle, beskriv oppbyggning av dataset (klasser og antall bilder), se om det fins mer info om hvor data er produsert på kaggle.

A description of where the data comes from, how it was procured for this study and (if it comes from various sources), how it was combined into a single data set.

1. **Interesting insights**

* Mange likheter på data, som senere byr på vanskligheter med klassifisering

An overview of any insights which may be gained from simply analysing the data by manual scrutiny or using tools such as SQL and/or functions in a spreadsheet editor.

1. **Data pre-processing**

* Beskrivelse av dataprocessor klasse, loading av data, forbereding av data (train, val test) augmentering av data. Skalering og flatten data slik at den passer ML, summering av rader og kolonner, grafer.
* Hvorfor vi lagde egen klasse istedenfor å bruke eksisterende library

Which types of data pre-processing was applied to the data and why? Provide code examples of how the pre-processing was done.

1. **Measuring performance**

* precision recall f1-score support
* Accuracy

Which metrics were used to assess any models developed in the study and why these specific one(s) were selected?

1. **Algorithms applied**

An overview of at least 3 different types of machine learning algorithms that have been applied to the data. These algorithms may be applied for the same purpose, i.e., for the sake of comparison to select the best performing algorithm or they may be applied for different reasons, i.e., to gather different insights.

For each machine learning algorithm provide an overview of the algorithm, why it was selected and how it performed. For all algorithms, provide code examples of how the models were created, trained, and tested/validated. Comparative results, such as outputs from different contrasting metrics, should be presented in tables.

* Classifikasjonsrapport, training tid, prediction tid

Sub-headings help structure your report document and make it easier for the reader to follow. It is advised that you create sub-headings for the (at least) 3 algorithms (models) which you create, as well as a separate one to compare and/or contrast their performance.

* Hvordan behandles bildene før bruk
* Gå I dybden på resultat fra KNN

**9.1 Algorithm 1**

**\* KNN**

A discussion of algorithm 1.

* Sammenligne resultater på (graf, augmenter og original)

**9.2 Algorithm 2**

**\* CNN**

A discussion of algorithm 2.

* Sammenligne original og aumented

**9.3 Algorithm 3**

**\* SVM (svc)**

A discussion of algorithm 3.

* Sammenligne resultater på (graf, augmenter og original)

**9.4 Comparison**

**\*** Training og prediction tid. Sammenligne accuracy og manuell testing.

\* Sammenligne resultater på alle modeler og forklar årsaken til at noen noen av modellene var bedre på bilder enn på graf og motsatt.

\*

A comparison of all three algorithms.

1. **Discussion (ca. 2300-2500 ord)**

Frank:

* Overraskende gode resultater på sum av rader/kolonner, gir rask prediction og training.
* Hadde resultater vært annerledes med større datasett, med mer variasjon? Lærer modellen kroppsdeler i stedet for sykdom?
* Vår opplevelse av de ulike modellene, hvordan var de å jobbe med.
* Størrelse på pickle, og årsak til dette (var ikke bruk så mye på grunn av enorme filstørrelser), raskere å generere enn å laste ned.
* Gridsearchcv på augmenterte bilder (SVM) kunne vært relevant å traine i lang tid for å få godt resultat, hvis dette hadde vært et oppdrag i realiteten.
* Vi ser at augmentering har stor påvirkning på resultatene, kunne augmentering vært annerledes for å få enda bedre resultater?

Gabriel:

* Etikk eller problemer med bruk i virkeligheten. Hva skjer om pasient får feil behandling på grunn av feil klassifisering.
* Hvor høy prosent klarer leger å diagnosere disse hudsykdommene? Kan det sammenlignes med resultatene vi fikk i modellene?
* Kan leger være bias? Noen sykdommer er mer skjelden og at legen kan anta at det er en mindre skjelden sykdom.
* Kan modellene brukes som bekreftelse på legens mistanke.

A discussion on what findings can be derived from the created models or a description of what utility the model(s) may hold when implemented in the real world.

1. **Conclusion (ca. 500-700 ord)**

**Frank:**

* CNN krever høy kompetanse, men har vært interresant og lærerikt å se hvordan de fungere.
* Overraskende gode resultater på sum av rader/kolonner, gir rask prediction og training.
* Lærerikt og veldig nyttig med DataProcessor og bruk av class for gjenbruk av kode

**Gabriel**

* For mye likheter I klassene gir problemer med klassifisering
* Er resultatene gode nok for faktisk bruk i virkeligheten?
* Referere til objective og om vi «oppnådde» målet

A conclusion in which you discuss why you feel that the objective(s) of the study has been met (or not). In the conclusion you also need to reflect on you journey throughout the project and any lessons you have learnt.

**References**

Add your reference list here.