

Business Understanding Report

Ammar Hasan 150454388

Report Introduction

This report documents the Business Understanding stage of the CRISP-DM process followed by this project. This stage documents decisions from a business/stakeholder perspective, highlighting the objectives and requirements of the data mining project to form its problem definition (the specification).

Organisation and Organisational Objectives

Even though this project is not directly related to any business or an organisation, this project needs to have a beneficial impact to problems faced by stakeholders that were not tackled before.

Background

This project objective is to develop machine learning model for the diagnosis of pigmented skin lesions, and to evaluate the models using various model evaluation techniques (confusion matrices, F1 scores, etc.).

Resources and Data

The dataset provided is based on a dataset from a kaggle competition. The dataset is called the HAM10000 ("Human Against Machine with 10000 training images") dataset, which is a set of dermoscopic images collected from various populations. The final dataset consists of 10015 images.

The ground truths are confirmed through histopathology follow-up examination (follow_up), expert consensus (consensus), or by in-vivo confocal microscopy (confocal).

Success Criteria and Objectives

To obtain what criterion should the models pass a look into what stakeholders on the field expect. To do this, a look into published work in the area is necessary.

Project Specification and Plan

Requirements and Rationale

Tools

Project Plan and Risks