HAN master applied data science project description

Author	Gerben Heinen
Date	20 Dec 2024
Document status	DRAFT

Context

Sanovo Technology Netherlands develops, produces, and sells egg grading machine in an international market. Our customers are located all over the world, in Europe, North-America, South-America, Africa, Asia and Oceania. In our office in the Netherlands all software, electrical and mechanical development are done by a strong team of engineers with a multidisciplinary background. More information can be found on www.sanovogroup.com[1].

Performance and maintenance data is collected from our grading machines for more than 2 years in our Linkpro platform based on the solution of Ixon[2]. This data is used to monitor the performance of our machine, where performance is determined by the amount of eggs that are graded in a specific time period. Obviously a grader operator would like to be able to grade as much eggs as possible in a shorter time.

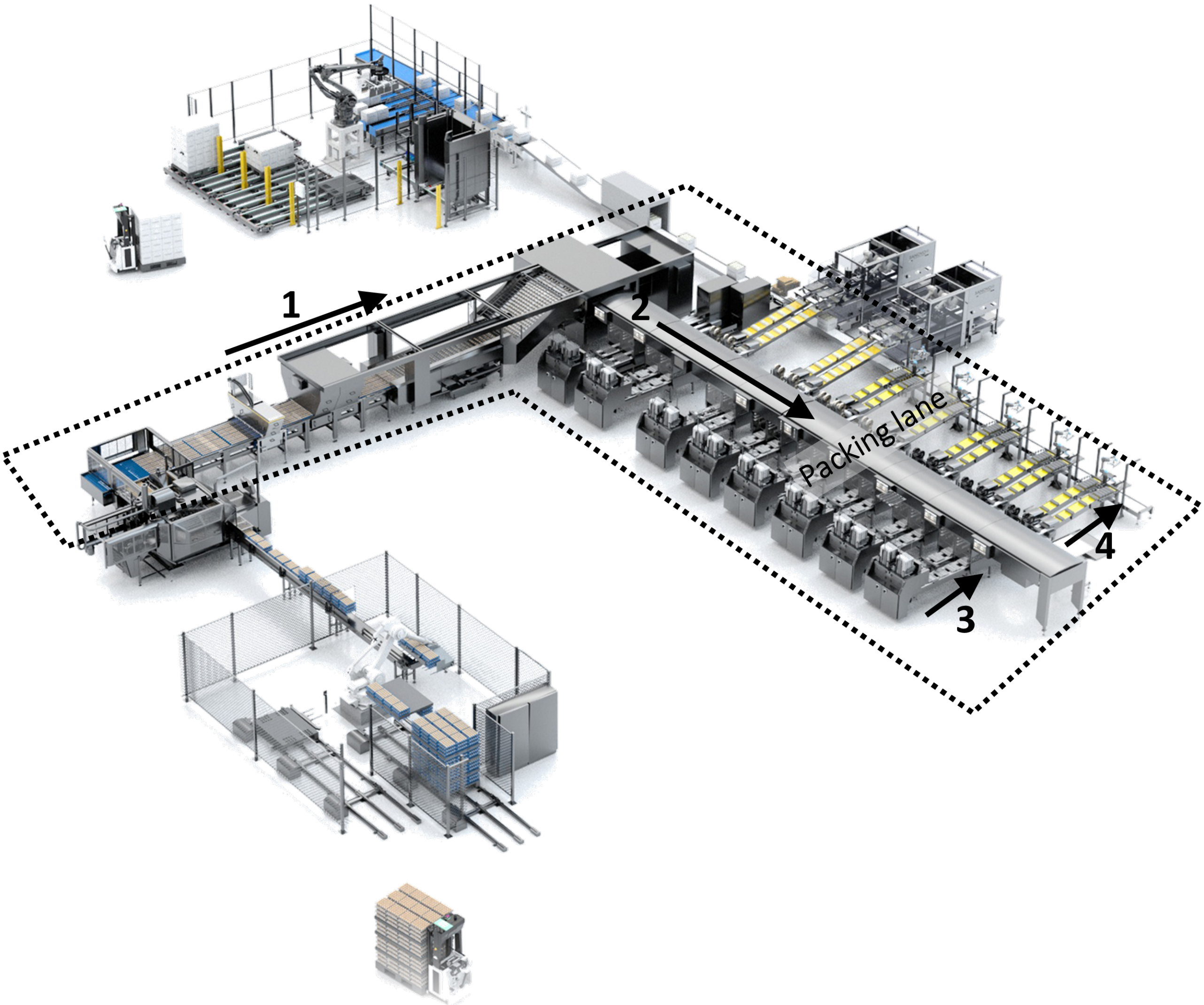
This document describes in an overview how an egg grader works, and what the problem description for this assignment is.

Egg grader

The image below shows an example of an egg grader. The dashed lines enclose the grader, the equipment outside the dashed lines are robotic solutions to get unsorted eggs on the grader and to pack the egg packs into cases and eventually on pallets.

The arrows show the basic flow of eggs through an egg grader.

1 http://www.sanovogroup.com/
2 https://www.ixon.cloud/en/



1. Eggs are transported into the grader on the infeed. In this part the eggs are weighed and their quality is measured with e.g. vision systems. The quality of an egg is determined by the presence of dirt, cracks, missing shell and blood inside the egg.
2. Eggs are transported to the packing lanes with the main track. The main track drops the eggs on the packing lanes. What packing lane is determined by how the grader operator has programmed the grader. For example eggs with Large size can be dropped on packing lane 1, and eggs with Medium size can be dropped on packing lane 2.
3. A packing lane contains the empty egg packs into which the main track drops the eggs. At position 3 an empty pack is 'denested' and put on the packing lane.
4. At position 4 eggs are in the pack and the pack is closed and transported further for being packed into a case.

Problem description

The performance of the grader in total is besides other factors, also determined by how well the packing lanes function. If a packing lane has a lot of errors, it will stop working temporarily and as a consequence the grader's main track can not drop the eggs on it. If enough packing lanes stop working than the grader is not able to drop eggs at all and the machine has to stop too which results in performance loss.

> Issues in the infeed or in the main track itself could also be cause for the grader to stop, but those are out of scope for this project.

A packing lane can stop because of an error on it, some examples of errors are:

- denester error
- missing egg error
- double pack error
- closer error

These errors are logged in our linkpro platform of each packing lane individually.

The following machine parameters are also logged:

- amount of re-routes: a re-route happens when the machine can not drop an egg on a packing lane and has to send it to another packing lane
- amount of machine stops: the machine stops when no packing lane is available that can take the egg
- amount of sorted eggs.

An example graph of this data is shown below.

So logically an increase in lane errors should also result in an increase in machine stops, but the relationship is extremely difficult to spot in the data.

Assignment

The assignment is to find the relationship between lane errors and an increase in machine stops in the data and to develop a model that can predict the number of machine stops on the basis of lane errors.

Appendix - data

Exported data of graderpro machines is available. The graderpro comes in different capacities which is measured in “cases per hour”, where a case is 360 eggs. So GP220 is a graderpro 220, GP400 is a graderpro 400, etc. The last 5 digits of the data file name is the project number, eg. 'han_data_science_info_gp75_18650.csv' is the graderpro 75 of customer 18650.

Data is in csv format. Headers explain the data:

- CASES_PROCESSED = amount of cases processed by the machine
- CECx = eggs graded of packing lane x.
- DECx = denester errors of packing lane x
- MECx = missing eggs of packing lane x
- CLECx = closer errors of packing lane x
- STOPS = machine stops
- REROUTES = machine reroutes