***ANALYSIS OF DELIVERY TIME PREDICTIONS***

*Natalia Wrześniak*

# Abstract

# Main goals

One of the aim of this report is to show how naive and incorrect current algorithm of predicting delivery time is. Now it calculates the mean from all collected data and applies it to every future order. It seams that there has to be better method which takes into consideration more variables available in used database. And this is the second main goal of this report - explore correlations or trends that could be valuable for prediction quality improvement.

# Assumptions and methodology

# Analysis

1. Generated a histograms showing the actual delivery length with 1 minute granularity.

Obraz zawierający tekst, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznieObraz zawierający tekst, zrzut ekranu, Wykres, diagram

Opis wygenerowany automatycznie

Chart 2 Distribution of actual delivery length while length is less than 15 min

Chart 1 Distribution of actual delivery length for all delivered orders

I took into consideration only orders with delivery length larger than 0 (some of the records in database have mistakes and end time is before start time).

For most orders the delivery length is between 0 and 11 minutes [Chart 2]

1. Generated a histogram showing prediction error (difference between planned and

actual delivery times).

Obraz zawierający tekst, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, Prostokąt

Opis wygenerowany automatycznie

Chart 4 Distribution of prediction error [in %]

Chart 3 Distribution of prediction error[ in seconds]

Two rows were excluded from analysis because of the huge error.

For orders of which both planned time and delivery time are know I calculated ‘prediction error’ as planned delivery time minus actual delivery time. As shown on [Chart 3] planned time is in most cases longer than actual. The error expressed in seconds varies between 0 to 10 minutes.

On [Chart 4] I presented distribution of predicted error in percent calculated by dividing absolute value of previously counted error by actual delivery time. Percentage error varies between 0 and 70%, which is a lot and something should be changed in the way of counting predicted time.

1. Visualized drivers’ hypothesis that that delivering in one of the sectors is significantly longer than in other sectors.

Obraz zawierający tekst, zrzut ekranu, Prostokąt, diagram

Opis wygenerowany automatycznieObraz zawierający tekst, zrzut ekranu, Prostokąt, fioletowy

Opis wygenerowany automatycznie

Chart 5 Mean Actual Delivery Duration by Sector

Chart 6 Mean Planned Delivery Duration by Sector

The hypothesis is true. It is visually proven on [Chart 5]. The mean actual delivery time in sector ‘2’ is more than two times longer than for the other two.

Also on the [Chart 6] I presented bar plot which in comparison to [Chart 5] presents how inaccurate current method of predicting delivery time is. For example the mean value of delivery time in sector “2” is around 800 seconds based on real data, but mean predicted value is four times less.

1. Checked correlations between order weight and actual delivery length.

There seems to be no correlation between total order’s weight and actual delivery length. But it is worth noting that for this comparison I could only use sixty four rows (only these rows contain all issues of interest).

1. Checked correlations between driver performance and actual delivery length.

There is an existing correlation between driver performance and actual delivery time. As shown in [Chart 8] drivers with ID ‘1’ or ‘3’ delivers packages the longest. It is extremely important while delivering orders in sector ‘2’ [Chart 9] where delivery times for driver '1' are more than three times longer and for driver '3' are two and a half times longer compared to the other two drivers.

1. Checked correlations between delivery time and actual delivery length.

# Conclusions