

Masterscriptie Business Analytics

---

# Op tijd vertrekken

Verklaringen voor ADC dispunctualiteit op ICA vluchten

---

Nivard van Wijk

**Begeleiders:**

dr. Evert Haasdijk  
drs. Simone van Neerven  
dr. Fetsje Moné-Bijma

Juli 2013

PUBLIEKE VERSIE



KLM Tactical Planning  
Havenmeesterweg 1  
1118 BG Luchthaven Schiphol



Vrije Universiteit Amsterdam  
Faculteit der Exacte Wetenschappen  
De Boelelaan 1081a  
1081 HV Amsterdam



# Voorwoord

De masteropleiding Business Analytics aan de Vrije Universiteit van Amsterdam wordt afgesloten met een afstudeerstage. Het doel van deze stage is om de verschillende onderdelen van de opleiding, namelijk bedrijfskunde, informatica en wiskunde, terug te laten komen in een onderzoek naar een specifiek bedrijfsprobleem. Deze masterscriptie is geschreven in opdracht van de business unit Hub Operations van de KLM.

In eerste instantie wil ik graag de KLM bedanken voor de mogelijkheid die zij mij hebben geboden om mijn afstudeeronderzoek uit te voeren. In het bijzonder gaat mijn dank uit naar mijn begeleidster Simone van Neerven en de (ex-)medewerkers van de afdeling Tactical Analysis: Ben Versteeg, Serdar Cifoglu, Ineke Robertus, Atilla Selçuk en Richard Hofman, voor de vrijheid en adviezen die zij tijdens mijn onderzoek hebben gegeven.

Vanuit de VU wil ik graag mijn begeleider Evert Haasdijk bedanken. Zijn aanwijzingen en enthousiasme zijn van zeer groot belang geweest voor het slagen van dit onderzoek. Verder gaat mijn dank uit naar Fetsje Moné-Bijma, die de taak van tweede lezer op haar heeft genomen. Mijn mede-studenten en in het bijzonder Nick Groen, wil ik danken voor het meedenken bij alle problemen die tijdens dit onderzoek opkwamen.

Nivard van Wijk  
- Amsterdam, 2013



# Inhoudsopgave

<b>1</b>	<b>Inleiding</b>	<b>1</b>
1.1	Probleemstelling . . . . .	1
1.2	Literatuuronderzoek . . . . .	2
1.2.1	Wetenschappelijk onderzoek . . . . .	2
1.2.2	Intern KLM onderzoek . . . . .	3
1.3	Structuur . . . . .	3
<b>2</b>	<b>Afhandelingsproces van vluchten</b>	<b>5</b>
2.1	Type vluchten . . . . .	5
2.2	Losse aankomst . . . . .	6
2.3	Los vertrek . . . . .	7
2.4	Omdraai . . . . .	7
<b>3</b>	<b>Data beschrijving</b>	<b>9</b>
3.1	Data voorbereiding . . . . .	9
3.2	Vluchteigenschappen . . . . .	10
3.3	Kritieke subprocessen . . . . .	10
3.4	Verband ADC en kritieke subprocessen . . . . .	12
<b>4</b>	<b>Data mining methoden</b>	<b>13</b>
4.1	Performance meting . . . . .	13
4.2	Decision tree . . . . .	14
4.2.1	C4.5 algoritme . . . . .	15
4.2.2	Ongebalanceerde data in C4.5 . . . . .	15
4.3	Subgroup discovery . . . . .	16
4.3.1	Pattern team . . . . .	17
4.3.2	Verschil tussen subgroup discovery en decision trees . . . . .	17
4.4	Support vector machines . . . . .	18
4.4.1	Soft-margin SVM . . . . .	18
4.4.2	Non-linear SVM . . . . .	19
4.4.3	Ongebalanceerde data in SVM . . . . .	20
4.4.4	Belangrijkste variabelen . . . . .	20
4.4.5	Versnellen van het SVM-RFE algoritme . . . . .	20
4.5	Implementatie . . . . .	21
4.5.1	J48 in WEKA . . . . .	21
4.5.2	SD in Cortana . . . . .	22

4.5.3 Libsvm . . . . .	22
<b>5 Resultaten</b>	<b>23</b>
5.1 Decision tree met resampling . . . . .	23
5.2 Decision tree unsampled . . . . .	24
5.3 Subgroup discovery . . . . .	25
5.4 Support vector machines . . . . .	26
5.5 Vergelijking boarding patronen . . . . .	26
5.6 Vergelijking data mining methoden . . . . .	27
<b>6 Conclusie</b>	<b>29</b>
6.1 Discussie . . . . .	30
6.2 Aanbevelingen . . . . .	31
<b>A Empirische verdeling doorlooptijd</b>	<b>33</b>
<b>B Variabelen per vlucht</b>	<b>35</b>
<b>C Constructie van SVM</b>	<b>39</b>
C.1 Optimale hypervlak methode . . . . .	39
C.2 Soft-margin SVM methode . . . . .	40
C.3 Non-linear SVM methode . . . . .	41
C.4 Kernel functies . . . . .	42
<b>D Aangepast SVM-RFE algoritme</b>	<b>43</b>
<b>E Beam-search Subgroup Discovery</b>	<b>45</b>
<b>F Gevonden patronen</b>	<b>47</b>
<b>G Belangrijkste variabelen per subproces</b>	<b>51</b>
<b>Literatuurlijst</b>	<b>53</b>

# Literatuurlijst

- [1] Alpaydin, E.: *Introduction to Machine Learning*. The MIT Press, 2010, ISBN 978-0-262-01243-0.
- [2] Atzmüller, M.: *Knowledge-intensive Subgroup Mining: Techniques for Automatic and Interactive Discovery*. proefschrift, Bayerischen Julius-Maximilians-Universität Würzburg, 2006.
- [3] Boser, B., I. Guyon, en V. Vapnik: *A training algorithm for optimal margin classifiers*. In *Proceedings of the fifth annual workshop on Computational learning theory*, pag. 144–152. ACM, 1992.
- [4] Chawla, N. V.: *C4.5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure*. In *Proceedings of the ICML'03 Workshop on Class Imbalances*, 2003.
- [5] Cortes, C. en V. Vapnik: *Support-vector networks*. *Machine learning*, 20(3):273–297, 1995.
- [6] Estabrooks, A., T. Jo, en N. Japkowicz: *A multiple resampling method for learning from imbalanced data sets*. *Computational Intelligence*, 20(1):18–36, 2004.
- [7] Gamberger, D. en N. Lavrač: *Expert-Guided Subgroup Discovery: Methodology and Application*. *Journal of Artificial Intelligence Research*, 17:501–527, 2002.
- [8] Ghattas, B. en A. B. Ishak: *An efficient method for variable selection using SVM based criteria*. Preprint IML, 2007.
- [9] Hsu, C. W., C. C. Chang, C. J. Lin, *et al.*: *A practical guide to support vector classification*, 2003.
- [10] Knobbe, A. J. en E. K. Ho: *Pattern teams*. In *Knowledge Discovery in Databases: PKDD 2006*, pag. 577–584. Springer, 2006.
- [11] Kruskal, W. H. en W. A. Wallis: *Use of ranks in one-criterion variance analysis*. *Journal of the American statistical Association*, 47(260):583–621, 1952.
- [12] Liu, X., J. Beltran, N. Mohanchandra, en G. Toussaint: *On speeding up support vector machines: proximity graphs versus random sampling for pre-selection condensation*. *World Academy of Science, Engineering and Technology*, 73:905 – 912, 2013.
- [13] Mitchell, T.: *Machine learning*. McGraw-Hill, 1997, ISBN 0-07-115467-1.
- [14] Mueller, E. R. en G. B. Chatterji: *Analysis of aircraft arrival and departure delay characteristics*. In *Proceedings of the AIAA Aircraft Technology, Integration, and Operations (ATIO) Conference, Los Angeles, CA*. Citeseer, 2002.
- [15] Neerven, S. van: *A0 effecten op hub SPL: NOC, MHB, ADC*. KLM intern, juli 2012.

- [16] Neerven, S. van: *Boarding proces irt PDC: Nadere analyse tbv targetsetting*. KLM intern, januari 2013.
- [17] Osuna, E., R. Freund, en F. Girosi: *Support vector machines: Training and applications*. A.I. Memo, (1602), 1997.
- [18] Quinlan, J.R.: *C4.5: programs for machine learning*. Morgan kaufmann, 1e druk, 1993, ISBN 1-55860-238-0.
- [19] Rakotomamonjy, A.: *Variable selection using svm based criteria*. The Journal of Machine Learning Research, 3:1357–1370, 2003.
- [20] Selçuk, A.: *ADC-performance: EUR 73J-73H-73W*. KLM intern.
- [21] Vapnik, V.: *The nature of statistical learning theory*. Springer, 1995, ISBN 0-387-98780-0.
- [22] Veropoulos, K., C. Campbell, en N. Cristianini: *Controlling the Sensitivity of Support Vector Machines*. In *Proceedings of the International Joint Conference on AI*, pag. 55–60, 1999.
- [23] Wang, J., P. Neskovic, en L. N. Cooper: *Selecting data for fast support vector machines training*. In *Trends in neural computation*, pag. 61–84. Springer, 2007.
- [24] Wilcoxon, F.: *Individual comparisons by ranking methods*. Biometrics bulletin, 1(6):80–83, 1945.
- [25] Wu, C. L.: *Inherent delays and operational reliability of airline schedules*. Journal of Air Transport Management, 11(4):273–282, 2005.
- [26] Wu, C. L. en R. E. Caves: *Aircraft operational costs and turnaround efficiency at airports*. Journal of Air Transport Management, 6(4):201–208, 2000.
- [27] Wu, C. L. en R. E. Caves: *Flight schedule punctuality control and management: A stochastic approach*. Transportation planning and technology, 26(4):313–330, 2003.
- [28] Wu, C. L. en R. E. Caves: *Modelling and simulation of aircraft turnaround operations at airports*. Transportation Planning and Technology, 27(1):25–46, 2004.