Chicago Air Quality

Hannah Lyon, Alan Flint, Michael Schulze,
Samarth Inani, Rushil Sheth

What went wrong?

Flight Delays

Hannah Lyon, Alan Flint, Michael Schulze,
Samarth Inani, Rushil Sheth

Customer service & reservations
BRITISH AIRWAYS

Analytic Goals

Our analytic goal was to see if flight delay times could be predicted using factors like airline, location, and time of year.

DATA DESCRIPTION

DAY OF WEEK

MONTH

YEAR

DISTANCE

SCHEDULED DEPARTURE

SCHEDULED ELAPSED TIME

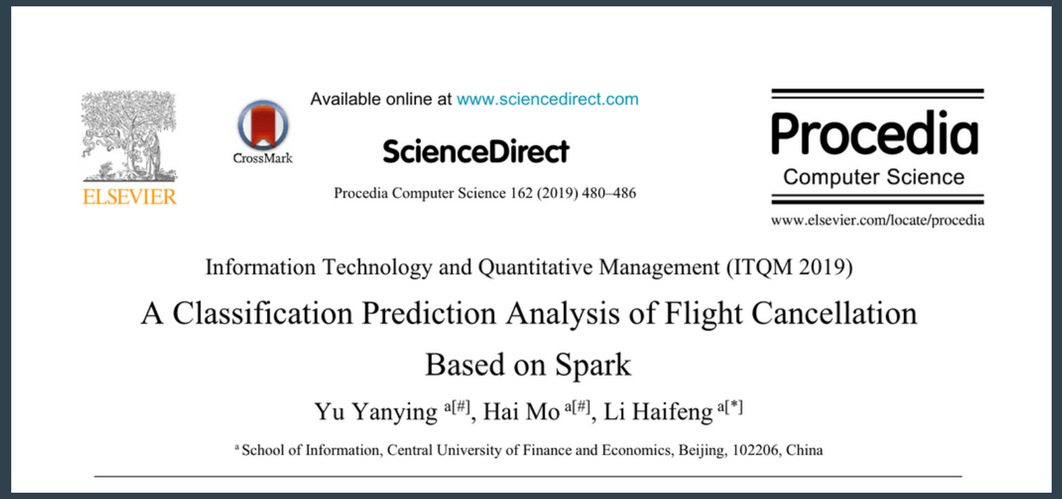
CARRIER

DESTINATION

ORIGIN

Pre-processing time 92.45 seconds on m5d.12xlarge

RELATED WORK



ELSEVIER

CrossMark

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 162 (2019) 480–486

Procedia Computer Science

www.elsevier.com/locate/procedia

Information Technology and Quantitative Management (ITQM 2019)

A Classification Prediction Analysis of Flight Cancellation Based on Spark

Yu Yanying [a][#], Hai Mo [a][#], Li Haifeng [a][*]

[a]School of Information, Central University of Finance and Economics, Beijing, 102206, China

International Journal of Engineering Research And Management (IJERM)
(An ISO 9001:2008 certified International, Peer reviewed, Open Access Journal, Approved by Govt. of Rajasthan, India)
ISSN: 2349- 2058, Volume-06, Issue-09, September 2019

Prediction of Flight Delays and Cancellation

Keerthana K, Viswa Harini T N, Deivarani S

FLIGHT CLASSIFICATION

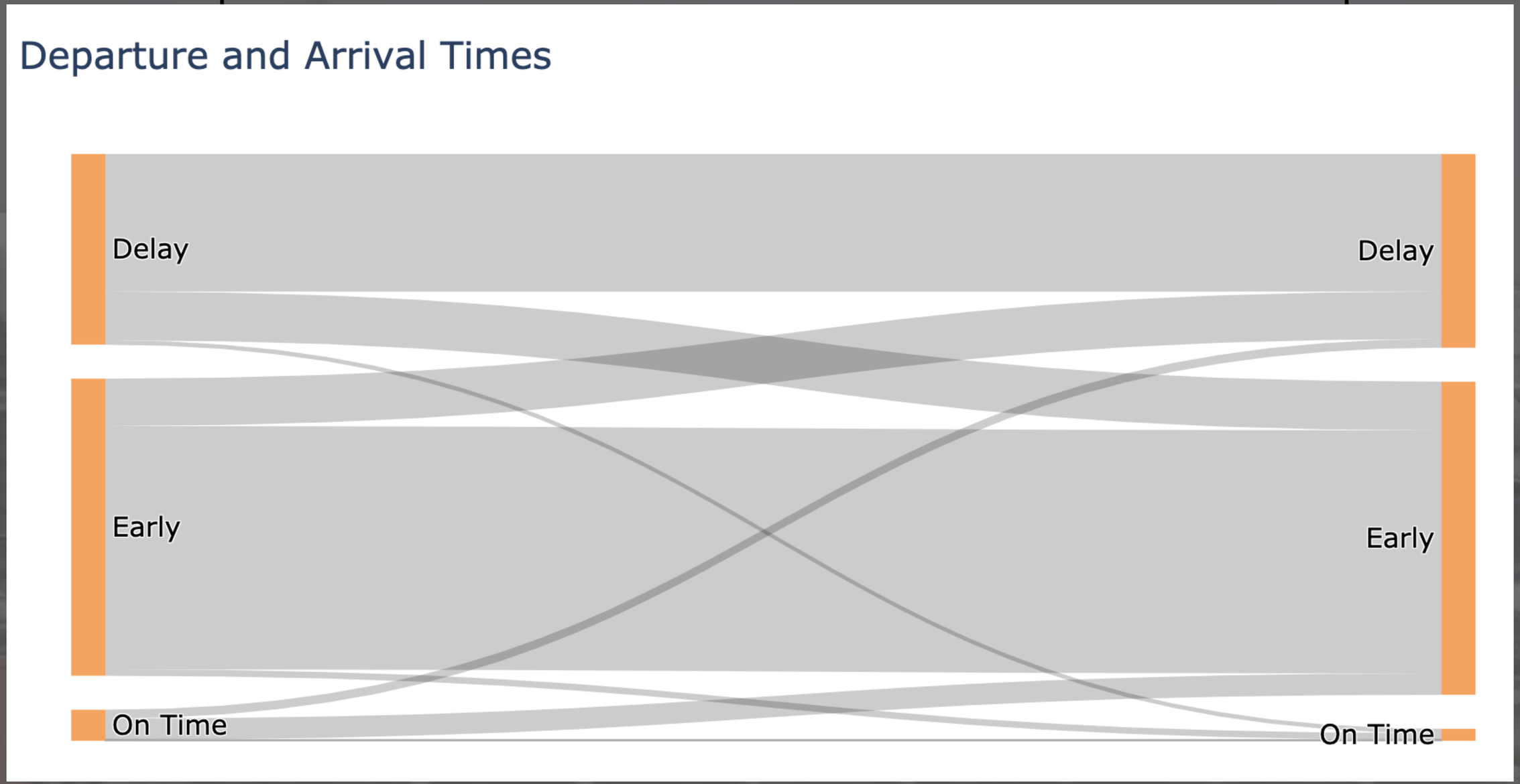
Uses Spark to run different classification algorithms on different numbers of nodes

DELAY PREDICTION

Aims to find the most consistently on time airline using an ANOVA model

EDA

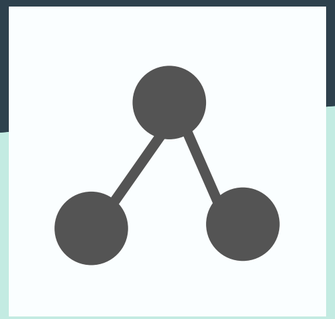
Departure and Arrival Flows



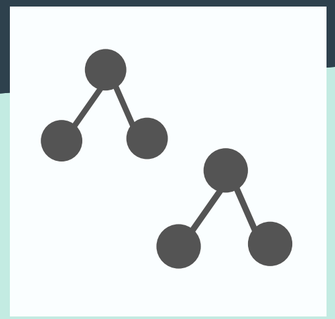
Models

(Ridge) Linear Regression

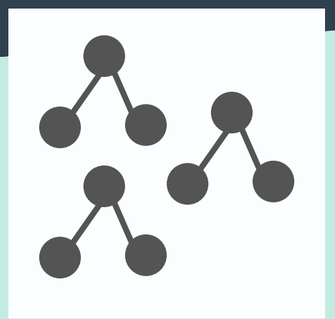
(Lasso) Linear Regression



Decision Tree



Random Forest



GBT

RMSE Scores

CONFIGURATION:
M5.16XLARGE
1 MASTER, 4 CORES

LINEAR REGRESSION(R)	12.78
LINEAR REGRESSION(L)	12.83
DECISION TREE	13.18
GRADIENT BOOSTED TREE	21.57
RANDOM FOREST	23.68

Best Model

Configurations

SPECS	c5d.18xlarge	m5.8xlarge	m5.16xlarge	m5.12xlarge
Master	1	1	1	1
Core	4	3	3	2
vCore	72	32	64	48
RAM(GB)	144	128	256	192
Storage(GB)	1800	512	1024	768
Time(s)	207	405	240	446

Key Takeaways

1. OUR MODEL IS ABLE TO PREDICT WITH AN RMSE OF UNDER 13
2. SOMETIMES SIMPLER MODELS ARE BETTER
3 PIVOT!!!! - AIRLINES, DATA FORMAT, DATASETS, LIFE
4. DON'T FLY WITH SAMARTH

THANK YOU!