APPLICATION OF DATA IN AVIATION INDUSTRY IDEA PROPOSAL

by

Jacinth Daniel Moses

Submitted for the degree of

BSc Computer Science

HERIOT-WATT UNIVERSITY
DUBAI

September 2023

Contents

Chapter 1

Airlines Delay and Cancellation Prediction Optimization for DXB

1.1 Problem Space

- Plane delays and cancellations are a huge inconvenience, not only for the passengers, crew and pilots flying but also for the airline who incurs financial losses due tp the delays and the airport who were expecting the resources allocated to the plane to be free.
- A solution where delays could be detected ahead of time by all parties can help with planning that while may not be able to prevent it, can minimize the cost financial or otherwise to all parties.
- However current prediction software is not verry accurate as it is inherantly difficult to gather data that correlates to delays or cancellations.

1.2 Data Sets

- Jared Brooks Kaggle Datset
- Gabriel Atkin Kaggle Datset
- FabienDaniel Kaggle Datset

1.3 Publications

- Airline Flight Delay Prediction Using Machine Learning Models
- Analysis of Flight Delay and Cancellation Prediction Based on Machine Learning Models
- A Classification Prediction Analysis of Flight Cancellation Based on Spark
- Flight delay/cancellation prediction using machine learning Adapting new ways to help stranded passengers

1.4 Hypothesis

- Most of the Research as well as data set links above are 2+ years old. Using newer machine learning models and with the current abundance of data we could not only produce more accurate results but could potentially produce them faster

NO.	Field Type	Length (bytes)	Description
1	IN_BYTES	N*4	Incoming counter with length N x 8 bits for number of bytes associated with an IP Flow.
2	IN_PKTS	N*4	Incoming counter with length N x 8 bits for the number of packets associated with an IP Flow
3	FLOWS	N	Number of flows that were aggregated; default for N is 4
4	PROTOCOL	1	IP protocol byte
5	SRC_TOS	1	Type of Service byte setting when entering incoming interface
6	TCP_FLAGS	1	Cumulative of all the TCP flags seen for this flow
7	L4_SRC_PORT	2	TCP/UDP source port number i.e.: FTP, Telnet, or equivalent
8	IPV4_SRC_ADDR	4	IPv4 source address
9	SRC_MASK	1	The number of contiguous bits in the source address subnet mask i.e.: the submask in slash notation
10	INPUT_SNMP	N	Input interface index; default for N is 2 but higher values could be used
11	L4_DST_PORT	2	TCP/UDP destination port number i.e.: FTP, Telnet, or equivalent
12	IPV4_DST_ADDR	4	IPv4 destination address
13	DST_MASK	1	The number of contiguous bits in the destination address subnet mask i.e.: the submask in slash notation
14	OUTPUT_SNMP	N	Output interface index; default for N is 2 but higher values could be used
15	IPV4_NEXT_HOP	4	IPv4 address of next-hop router
16	SRC_AS	N*2	Source BGP autonomous system number where N could be 2 or 4
17	DST_AS	N*2	Destination BGP autonomous system number where N could be 2 or 4
18	BGP_IPV4_NEXT_HOP	4	Next-hop router's IP in the BGP domain
19	MUL_DST_PKTS	N*4	IP multicast outgoing packet counter with length N x 8 bits for packets associated with the IP Flow
20	MUL_DST_BYTES	N*4	IP multicast outgoing byte counter with length N x 8 bits for bytes associated with the IP Flow
21	LAST_SWITCHED	4	System uptime at which the last packet of this flow was switched
22	FIRST_SWITCHED	4	System uptime at which the first packet of this flow was switched
23	OUT_BYTES	N*4	Outgoing counter with length N x 8 bits for the number of bytes associated with an IP Flow
24	OUT_PKTS	N*4	Outgoing counter with length N x 8 bits for the number of packets associated with an IP Flow.
25	MIN_PKT_LNGTH	2	Minimum IP packet length on incoming packets of the flow
26	MAX_PKT_LNGTH	2	Maximum IP packet length on incoming packets of the flow
27	IPV6_SRC_ADDR	16	IPv6 Source Address
28	IPV6_DST_ADDR	16	IPv6 Destination Address
29	IPV6_SRC_MASK	1	Length of the IPv6 source mask in contiguous bits
30	IPV6_DST_MASK	1	Length of the IPv6 destination mask in contiguous bits
31	IPV6_FLOW_LABEL	3	IPv6 flow label as per RFC 2460 definition
32	ICMP_TYPE	2	Internet Control Message Protocol (ICMP) packet type; reported as ((ICMP Type*256) + ICMP code)
33	MUL_IGMP_TYPE	1	Internet Group Management Protocol (IGMP) packet type
35	SAMPLING_ALGORITHM	1	The type of algorithm used for sampled NetFlow: 0x01 Deterministic Sampling ,0x02 Random Sampling
36	FLOW_ACTIVE_TIMEOUT	2	Timeout value (in seconds) for active flow entries in the NetFlow cache
37	FLOW_INACTIVE_TIMEOUT	2	Timeout value (in seconds) for inactive flow entries in the NetFlow cache
38	ENGINE_TYPE	1	Type of flow switching engine: RP = 0, VIP/Linecard = 1
39	ENGINE_ID	1	ID number of the flow switching engine
40	TOTAL_BYTES_EXP	N*4	Counter with length N x 8 bits for bytes for the number of bytes exported by the Observation Domain
41	TOTAL_PKTS_EXP	N*4	Counter with length N x 8 bits for bytes for the number of packets exported by the Observation Domain
42	TOTAL_FLOWS_EXP	N*4	Counter with length N x 8 bits for bytes for the number of flows exported by the Observation Domain
43	*Vendor Proprietary*		
44	IPV4_SRC_PREFIX	4	IPv4 source address prefix (specific for Catalyst architecture)
45	IPV4_DST_PREFIX	4	IPv4 destination address prefix (specific for Catalyst architecture)
46	MPLS_TOP_LABEL_TYPE	1	MPLS Top Label Type: 0x00 UNKNOWN 0x01 TE-MIDPT 0x02 ATOM 0x03 VPN 0x04 BGP 0x05 LDP
47	MPLS_TOP_LABEL_IP_ADDR	4	Forwarding Equivalent Class corresponding to the MPLS Top Label
48	FLOW_SAMPLER_ID	1	Identifier shown in "show flow-sampler"
49	FLOW_SAMPLER_MODE	1	The type of algorithm used for sampling data: 0x02 random sampling. Use in connection with FLOW_SAMPLER_MODE
50	FLOW_SAMPLER_RANDOM_INTERVAL	4	Packet interval at which to sample. Use in connection with FLOW_SAMPLER_MODE

51	*Vendor Proprietary*		
52	MIN_TTL	1	Minimum TTL on incoming packets of the flow
53	MAX_TTL	1	Maximum TTL on incoming packets of the flow
54	IPV4_IDENT	2	The IP v4 identification field
55	DST_TOS	1	Type of Service byte setting when exiting outgoing interface
56	IN_SRC_MAC	6	Incoming source MAC address
57	OUT_DST_MAC	6	Outgoing destination MAC address
58	SRC_VLAN	2	Virtual LAN identifier associated with ingress interface
59	DST_VLAN	2	Virtual LAN identifier associated with egress interface
60	IP_PROTOCOL_VERSION	1	Internet Protocol Version Set to 4 for IPv4, set to 6 for IPv6. If not present in the template, then version 4 is assumed.
61	DIRECTION	1	Flow direction: 0 - ingress flow, 1 - egress flow
62	IPV6_NEXT_HOP	16	IPv6 address of the next-hop router
63	BPG_IPV6_NEXT_HOP	16	Next-hop router in the BGP domain
64	IPV6_OPTION_HEADERS	4	Bit-encoded field identifying IPv6 option headers found in the flow
65 - 69	*Vendor Proprietary*		
70	MPLS_LABEL_1	3	MPLS label at position 1 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
71	MPLS_LABEL_2	3	MPLS label at position 2 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
72	MPLS_LABEL_3	3	MPLS label at position 3 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
73	MPLS_LABEL_4	3	MPLS label at position 4 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
74	MPLS_LABEL_5	3	MPLS label at position 5 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
75	MPLS_LABEL_6	3	MPLS label at position 6 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
76	MPLS_LABEL_7	3	MPLS label at position 7 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
77	MPLS_LABEL_8	3	MPLS label at position 8 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
78	MPLS_LABEL_9	3	MPLS label at position 9 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
79	MPLS_LABEL_10	3	MPLS label at position 10 in the stack. This comprises 20 bits of MPLS label, 3 EXP (experimental) bits and 1 S (end-of-stack) bit.
80	IN_DST_MAC	6	Incoming destination MAC address
81	OUT_SRC_MAC	6	Outgoing source MAC address
82	IF_NAME	N**	Shortened interface name i.e.: ”FE1/0”
83	IF_DESC	N**	Full interface name i.e.: ”’FastEthernet 1/0”
84	SAMPLER_NAME	N**	Name of the flow sampler
85	IN_ PERMANENT _BYTES	N*4	Running byte counter for a permanent flow
86	IN_ PERMANENT _PKTS	N*4	Running packet counter for a permanent flow
87	* Vendor Proprietary*		
88	FRAGMENT_OFFSET	2	The fragment-offset value from fragmented IP packets
105 to 127			**Reserved for future use by cisco**