บทที่ 4

ผลการทดลอง

การทดลองทำการตรวจสอบว่าตัวแปรต่างๆ มีความสัมพันธ์กันหรือไม่ปรากฏว่าได้ค่า KMO (Kaiser-Meyer-Olkin) = 0.851 การทดสอบ Bartlett's Test ให้ค่า Sig. =0.000 แสดงว่าข้อมูลมีความสัมพันธ์กัน ควรใช้เทคนิคการวิเคราะห์องค์ประกอบ (Factor Analysis) ได้ เมื่อสกัดปัจจัยและหมุนแกนปัจจัยสามารถแสดงรายละเอียดได้ดังตารางที่ 3

ตารางที่ 3: ตารางแสดงรายละเอียดปัจจัยที่ได้หลังหมุนแกนปัจจัย

ปัจจัยที่	ค่า Eigen value	ตัวแปรที่อยู่ในปัจจัยเดียวกันเรียงตามค่า PCA
1	5.68	V27, V28, V40,V41, V37
2	4.47	V35,V36,V1
3	3.15	V13,V16,V15,V14
4	2.215	V24,V23
5	2.067	V2
6	1.812	V29
7	1.619	V26,V25
8	1.393	V22,V10
9	1.333	V39,V38
10	1.182	V11,V9
11	1.035	V31
12	1.015	V17
13	1.003	V5,V18

จากตารางที่ 3 จะเห็นได้ว่าชุดข้อมูลดังกล่าวที่มีค่า Eigen value มากกว่า 1 สามารถสกัดได้เป็น 13 ปัจจัย ตัวแปรที่อยู่ในปัจจัยเดียวกันเรียงตามค่า PCA โดยเรียงจากค่ามากไปน้อย ผู้วิจัยมีความคิดว่าข้อมูลตัวแปรที่อยู่ในปัจจัยเดียวกันย่อมมีความคล้ายกันสูง ข้อมูลต่างปัจจัยกันย่อมมีความสัพพันธ์กันน้อย ดังนั้นผู้วิจัยจึงทำการเลือกตัวแปรตัวแรกของทุกปัจจัยมาเป็นคุณลักษณะเด่น ดังนั้นคุณลักษณะเด่นที่ได้รับการคัดเลือกมีทั้งหมด 13 ตัวแปรดังนี้ V27, V35, V13, V24, V2, V29, V26, V22, V39, V11, V31, V17, และ V5

ตารางที่ 4 : แสดงรายละเอียดคุณลักษณะของแต่ละวิธี

คุณลักษณะ	**รายละเอียดชื่อคุณลักษณะที่ได้รับเลือก**	**จำนวนคุณลักษณะ**
Full Features	duration, protocol_type, service, flag, src_bytes, dst_bytes, land, wrong_fragment, urgent, hot, num_failed_logins, logged_in, num_compromised, root_shell, su_attempted, num_root, num_file_creations, num_shells, num_access_files, num_outbound_cmds, is_host_login, is_guest_login, count, serv_count, serror_rate, srv_serror_rate, reerror_rate, srv_rerror_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate, dst_host_srv_rerror_rate,	41
GA approach	service, flag, wrong_fragment, hot, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_srv_serror_rate, duration, src_bytes, srv_serror_rate, land, urgent, su_attempted, num_root, num_shells, num_access_files, isguest_login	21
Kok approach	service, dst_bytes, logged_in, count, dst_host_count, root_shell, dst_host_rerror_rate	7
Factor Analysis approach วิธีที่นำเสนอ	reerror_rate, dst_host_diff_srv_rate, num_compromised, serv_count, protocol_type, same_srv_rate, srv_serror_rate, is_guest_login, dst_host_srv_serror_rate, num_failed_logins, srv_diff_host_rate, num_file_creations, src_bytes	13

ทำการสุ่มแบ่งชุดข้อมูลเป็น 3 ชุดมีรายละเอียดดังตารางที่ 5 ผลการทดลองแสดงใน ตารางที่ 6, 7, และ 8 ตามลำดับ

ตารางที่ 5: รายละเอียดของข้อมูล 3 ชุด

ชื่อชุดข้อมูล	**จำนวนข้อมูล (records)**
Dataset-1	186,745
Dataset-2	49,438
Dataset-3	25,419

ตารางที่ 6: ผลการทดลองกับชุด Dataset 1

ประเภทการบุกรุก	ค่าความถูกต้อง (Truth positive rate)			
	คุณลักษณะทั้งหมด	**วิธี GA**	**วิธีคุณ Kok**	**วิธีที่นำเสนอ**
Normal	99.97	99.95	99.81	99.96
Dos	99.96	97.21	99.39	99.96
Probing	97.83	100.00	78.26	100
R2L	12.03	6.01	0.00	12.08
U2R	15.38	12.82	23.08	23.08
ทุกประเภท	99.80	97.32	99.25	99.8

ตารางที่ 7: ผลการทดลองกับชุด Dataset 2

ประเภทการบุกรุก	ค่าความถูกต้อง (Truth positive rate)			
	คุณลักษณะทั้งหมด	**วิธี GA**	**วิธีคุณ Kok**	**วิธีที่นำเสนอ**
Normal	99.48	99.30	99.11	99.36
Dos	99.95	97.10	99.34	99.96
Probing	99.08	99.31	78.39	98.61
R2L	9.91	13.51	0.00	13.51
U2R	15.38	12.82	23.08	23.08
ทุกประเภท	99.58	97.30	98.82	99.57

ตารางที่ 8: ผลการทดลองกับชุด Dataset 3

ประเภทการบุกรุก	ค่าความถูกต้อง (Truth positive rate)			
	คุณลักษณะทั้งหมด	วิธี GA	วิธีคุณ Kok	วิธีที่นำเสนอ
Normal	99.70	99.44	99.22	99.66
Dos	99.96	97.28	99.38	99.94
Probing	99.67	99.33	78.00	99.33
R2L	15.00	5.00	0.00	20.00
U2R	15.38	12.82	23.08	23.08
ทุกประเภท	99.50	97.31	98.66	99.51

จากผลการทดลองจะเห็นได้ว่าค่าความถูกต้องมีค่าที่ใกล้เคียงกับเมื่อใช้คุณลักษณะทั้งหมด 41 ตัว ดังนั้นผลการทดลองยืนยันว่าสามารถใช้เทคนิคการวิเคราะห์องค์ประกอบมาคัดเลือกคุณลักษณะเด่นได้ ซึ่งขบวนการไม่ยุ่งยาก และไม่ซับซ้อน สามารถประมวลผลได้เร็วด้วยโปรแกรม SPSS