อนุสรา หิรัญวนากุล: เทคนิคการเพิ่มประสิทธิภาพการคาดการณ์ผลผลิตในอุตสาหกรรมการผลิตฮาร์ดดิสก์ไดรฟ์ (TECHNIQUES TO IMPROVE YIELD PREDICTION IN HARD DISK DRIVE MANUFACTURING INDUSTRY) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.นิตยา เกิดประสพ, 182 หน้า.

ในวงการอุตสาหกรรมการผลิตฮาร์ดดิสก์ไดรฟ์นั้นการคาดการณ์ผลผลิต (Yield Prediction) เป็นงานที่มีความสำคัญที่สุดอย่างหนึ่ง ความแม่นยำของการคาดการณ์ผลผลิตมีผลกระทบและบทบาทอย่างสูงในธุรกิจนี้ การคาดการณ์ผลผลิตให้มีความแม่นยำอาจจะเกิดอุปสรรคได้จากหลายปัจจัย โดยหนึ่งในปัจจัยสำคัญที่สุดคือการที่ข้อมูลการผลิตและการทดสอบผลิตภัณฑ์นั้นมีจำนวนมหาศาลทำให้มีความยากต่อการนำไปใช้งาน นอกเหนือจากนั้นแล้วข้อมูลเหล่านั้นมีความไม่สมดุลของข้อมูลระหว่างฮาร์ดดิสก์ไดรฟ์ที่ผ่านและไม่ผ่านการทดสอบอยู่ในระดับที่สูงมาก ด้วยเหตุนี้เองการสร้างโมเดลสำหรับการคาดการณ์ผลผลิตที่แม่นยำนั้นจึงเป็นเรื่องที่นับได้ว่ามีความท้าทายเป็นอย่างยิ่ง งานวิจัยนี้จึงได้นำเสนอการเพิ่มประสิทธิภาพในการทำงานของการคาดการณ์ผลผลิตในอุตสาหกรรมการผลิตฮาร์ดดิสก์ไดรฟ์โดยการนำความรู้ทางด้านการเรียนรู้ของเครื่องและวิธีการต่าง ๆ ทางด้านการทำเหมืองข้อมูล (Data Mining) เข้ามาช่วยจัดการปัญหา ทั้งการนำเทคนิคการทำข้อมูลให้สมดุล (Data Balancing) การนำอัลกอริทึมการเรียนรู้และวิธีการทางสถิติมาใช้ในการคัดเลือกคุณลักษณะจากผลลัพธ์ของการทดสอบผลิตภัณฑ์ รวมไปถึงการนำอัลกอริทึมการเรียนรู้มาใช้ในการสร้างโมเดลการคาดการณ์ผลผลิตของกระบวนการ นอกเหนือจากนั้นแล้วงานวิจัยนี้ยังได้นำเสนอเทคนิคใหม่คือการรวมกลุ่มข้อมูลมาประยุกต์ใช้ในการคำนวณผลผลิตโดยจะเป็นการรวมกลุ่มของข้อมูลในลักษณะของการแบ่งกลุ่มข้อมูลตามจำนวนค่าคงที่ที่ได้กำหนดไว้แทนที่วิธีการดั้งเดิมที่อ้างอิงการรวมกลุ่มของข้อมูลเป็นรายสัปดาห์เพื่อการคำนวณผลผลิต การทดสอบประสิทธิภาพใช้เกณฑ์การวัดค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) เป็นหลัก

จากผลการดำเนินงานวิจัยพบว่าเทคนิคใหม่ที่นำเสนอการทำสมดุลข้อมูลโดยใช้วิธีการ Data Balancing by k-Means Clustering k-Nearest Neighbors and Re-Sampling (DBC-2KAR) ซึ่งเป็นการประยุกต์อัลกอริทึมทางด้านการเรียนรู้ของเครื่องร่วมกับการสุ่มข้อมูลเข้าด้วยกันนั้นสามารถทำให้ข้อมูลเกิดความสมดุลและส่งผลให้โมเดลการเรียนรู้ในด้านการจำแนกสามารถคัดเลือกคุณลักษณะที่สำคัญออกมาได้ ขั้นตอนถัดมาในส่วนของการนำคุณลักษณะที่ได้มาดำเนินการสร้างโมเดลการคาดการณ์ผลผลิตด้วยอัลกอริทึม Artificial Neural Network และ Multiple Linear Regression ผนวกเข้ากับการรวมกลุ่มข้อมูลด้วยวิธีการรวมกลุ่มข้อมูลด้วยค่าคงที่

ด้วยวิธีการที่ได้กล่าวมาข้างต้นนั้นจะให้ผลลัพธ์ที่ดีกว่าวิธีการดั้งเดิมที่วิศวกรกระบวนการใช้งานอยู่ในปัจจุบัน โดยวิธีการที่ดีที่สุดได้แก่การคัดเลือกคุณลักษณะด้วยอัลกอริทึม Genetic Algorithm ร่วมกับการสร้างโมเดลการเรียนรู้ด้วยอัลกอริทึม Multiple Linear Regression โดยใช้การรวมกลุ่มแบบค่าคงที่ซึ่งจะให้ค่า MAE และ RMSE อยู่ที่ 0.559 และ 0.732 ตามลำดับ ถ้าหากเปรียบเทียบกับวิธีการดั้งเดิมแล้วค่าความผิดพลาดที่วัดได้จะลดลงไปถึง 60% โดยในส่วนของการทดลองการเพิ่มค่าคงที่ของการรวมกลุ่มข้อมูลนั้นพบว่าส่งผลให้มีค่าความผิดพลาดลดลงตั้งแต่ค่าคงที่เท่ากับ 1,000 และน้อยที่สุดเมื่อมีค่าเท่ากับ 10,000 จากนั้นแม้ว่าจะเพิ่มค่าคงที่การรวมกลุ่มข้อมูลไปจนถึง 40,000 ประสิทธิภาพในการคาดการณ์ผลผลิตจะอยู่ในระดับคงที่

มหาวิทยาลัยเทคโนโลยีสุรนารี

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2563 ลายมือชื่ออาจารย์ที่ปรึกษา_____

ANUSARA HIRUNYAWANAKUL : TECHNIQUES TO IMPROVE YIELD PREDICTION IN HARD DISK DRIVE MANUFACTURING INDUSTRY. THESIS ADVISOR : ASSOC. PROF. NITTAYAKERDPRASOP, Ph.D., 182 PP.

YIELD PREDICTION IN MANUFACTURING/MACHINE LEARNING/ IMBALANCED DATA/DATA BALANCING/FEATURE SELECTION/ GENETIC ALGORITHM/MULTIPLE LINEAR REGRESSION

In the field of hard disk drive manufacturing, yield prediction is one of the most important tasks. An accurate yield prediction has high contribution and influence on this business. There are many factors to obstruct the process of yield prediction modeling. One of key problems is tremendous amount of features collected from manufacturing and testing process. This problem creates difficulty for the data usage. Moreover, imbalance ratio between passed and failed hard disk drive units in the manufacturing data is extremely high. This issue makes the creation of an accurate yield prediction model a challenging task. This research introduces techniques to improve yield prediction performance in hard disk drive manufacturing by applying knowledge of machine learning and several techniques of data mining to solve the problem. Techniques of data balancing and the application of learning algorithm and statistics are used in data preparation and feature selection process. Machine learning techniques are also applied in modeling for yield prediction step. This research also introduces the novel method in data grouping step called data aggregation by consistency quantity to replace the original technique that data are grouped by week.

Performance has been evaluated through the Mean Square Error (MAE) and Root Mean Square Error (RMSE) measurements.

The experimental result shows that the proposed method, which is a combination of machine learning algorithms and resampling techniques called DBC-2KAR (Data Balancing k-Means Clustering, k-Nearest Neighbors and Resampling), is able to turn a highly imbalanced dataset to be a balanced and efficient one. Efficiency is due to the fact that the classification learning model can find the important attributes in feature selection step. These important attributes are to be used in the subsequent step for yield prediction modeling with algorithm Artificial Neural Network and Multiple Linear Regression combined with the novel technique called Data Aggregate by Fixed number that can provide better result than the traditional method adopted by current process engineers. The best methodology is the combination of feature selection using genetic algorithm and yield prediction modeling with multiple linear regression. RMSE and MAE of this combination are 0.559 and 0.732, respectively. In terms of error rate reduction from traditional method, this method can reduce the error rate by 60%. In the portion of experiment to increase fixed number of Data Aggregation, we found that errors of yield prediction are getting lower since the fixed number is 1,000. The lowest error of yield prediction is provided when fixed numbers is 10,000 and results are steady even if the fixed number has been increased up to 40,000.

School of Computer Engineering Student's Signature

Academic Year 2020 Advisor's Signature