ຄາດຄະເນໂອກາດເກີດພະຍາດຈາກຜົນການກວດເລືອດ ໂດຍນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນ

ສົມມິດ ທຸມມາລີ[1], ດາວເພັດ ວິໄລນາມ, ລັດສະໝີ ຈິດຕະວົງ

ພາກວິຊາ ວິທະຍາສາດຄອມພິວເຕີ, ຄະນະ ວິທະຍາສາດທຳມະຊາດ, ມະຫາວິທະຍາໄລແຫ່ງຊາດ, ສປປ ລາວ

[1]***ຕິດຕໍ່ພົວພັນ:*** *ສົມມິດ ທຸມມາລີ, ຄະນະ ວິທະຍາສາດທຳມະຊາດ, ເບີ ໂທ ແລະ WhatsApp: +856 20 5572 0450, Email: s.thoummaly@nuol.edu.la*

ຂໍ້ມູນບົດຄວາມ:

ການສົ່ງບົດ: 19 ຕຸລາ 2022

ການປັບປຸງ: 29 ມິຖຸນາ 2023

ການຕອບຮັບ: 07 ກໍລະກົດ 2023

ບົດຄັດຫຍໍ້

ໃນວິທະຍານິພົນສະບັບນີ້ມີຈຸດປະສົງ ເພື່ອສຶກສາຂັ້ນຕອນວິທີການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ໂດຍນຳໃຊ້ວິທີ ແລະ ເຕັກນິກການຈຳແນກຂໍ້ມູນໃນຮູບແບບທີ່ມີການຝຶກສອນເຊິ່ງເປັນໜຶ່ງໃນເຕັກນິກໃນການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ເຊິ່ງໄດ້ແບ່ງຂໍ້ມູນອອກເປັນສອງຊຸດຍ່ອຍຄື: ຊຸດຂໍ້ມູນການຝຶກສອນ ແລະ ຊຸດຂໍ້ມູນການທົດສອບ, ຊຸດຂໍ້ມູນການຝຶກສອນໃຊ້ເພື່ອສ້າງຮູບແບບ (Model) ສ່ວນຊຸດຂໍ້ມູນການທົດສອບໃຊ້ເພື່ອຫາປະສິດຕິພາບຂອງຮູບແບບ (Model) ໂດຍການປະເມີນປະສິດທິພາບຂອງຮູບແບບແມ່ນໃຊ້ມາຕຼິດສັບສົນ (Confusion Matrix) ເຊິ່ງຂໍ້ມູນ (Data Set) ໃນການຄົ້ນຄ້ວາຄັ້ງນີ້ ແມ່ນໄດ້ນຳໃຊ້ຜົນການກວດເລືອດໃນຊ່ວງປີ 2019-2020 ຂອງໂຮງໝໍ 5 ເມສາ ໃນຫ້ອງວິເຄາະວິທະຍາ ມາເປັນຂໍ້ມູນໃນການວິໄຈຄັ້ງນີ້ ເຊິ່ງໄດ້ນຳໃຊ້ວິທີການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ໃນສາມຮູບແບບຄື: ການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree), ເຄືອຂ່າຍປະສາດທຽມ (Neural Network), ການຈຳແນກແບບ Naive Bayes ແລະ ໃຊ້ເຄື່ອງມືໂປຣແກຣມ (Orange Tool) ໃນການຈຳແນກຜົນຂອງການຄົ້ນຄ້ວາ, ຫຼັງຈາກນັ້ນນຳເອົາຮູບແບບທີ່ດີທີ່ສຸດມາເປັນຕົວຕົ້ນແບບ (Model) ໃນການພັດທະນາໂປຼແກຼມ. ຈາກຜົນການທົດລອງສາມາດສະແດງຜົນທີ່ໄດ້ຮັບຄື: ຈາກຂໍ້ມູນ Data set 1543 ຄົນ, ທີ່ມີຄ່າຄວາມສ່ຽງ Risk 742 ຄົນ ແລະ ບໍ່ສ່ຽງ Non-Risk 801 ຄົນ, ເຊິ່ງໃນຮູບແບບການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree) ມີຄວາມຊັດເຈນ 97.1%, ໃນຮູບແບບເຄືອຂ່າຍປະສາດທຽມ (Neural Network) ມີຄວາມຊັດເຈນ 90.4% ແລະ ການຈຳແນກແບບ Naive Bayes ມີຄວາມຊັດເຈນ 88.2% ແລ້ວເອົາຮູບແບບ (Model) ທີ່ມີຄວາມຊັດເຈນສຸດຄື: ຮູບແບບການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree) ມາເປັນໂຕຕົ້ນແບບໃນການພັດທະນາໂປຼແກຼມ.

ຄຳສັບສຳຄັນ: *ການຂຸດຄົ້ນຂໍ້ມູນ, ການຕັດສິນໃຈແບບຕົ້ນໄມ້, ເຄືອຂ່າຍປະສາດທຽມ, Naive Bayes, ການກວດເລືອດ*

Diseases Hematology Prediction Using Data Mining Techniques

Sommith THOUMMLY[1], Daophet VILAINAM, Lathsamy CHIDTAVONG

Computer Science Department, Faculty of Natural Sciences, National University of Laos, Lao PDR

[1]Correspondence: *Sommith Thoummaly, Faculty of Natural Sciences, Tel & WhatsApp: +856 20 5572 0450, Email: s.thoummaly@nuol.edu.la*

Article Info:
Submitted: Oct 19, 2022
Revised: Jun 29, 2023
Accepted: Jul 07, 2023

Abstract

The purpose of this dissertation to study the procedures of data mining methods using methods and techniques in a trained format, which is part of the data mining techniques, which are divided into two sub-sets: The training data set and test data set, the training data set used to create the model while the test data set used to find the effectiveness of the model by evaluating the effectiveness of the model is using a confusion matrix The data in this research are based on the results of blood tests during the period 2019-2020 in the laboratory of Hamesa hospital as data in this research. It uses Data Mining in three ways: Decision Tree, Neural Network, Naive Bayes and Orange Tool to analyze of the research, then take the best model into the prototype in program development.

From the experimental results so the results can be as follows: From the data set of 1543 people, the risk value is 742 people and the non-risk person is 801 people. In the Decision Tree format with 97.1% accuracy, in the Neural Network with 90.4% and in the Naive Bayes mode with 88.2% accuracy, then take the model the most accurate is: Decision Tree as a model for program development.

Keywords: *Data Mining, Decision Tree, Neural Network, Naive Bayes, Complete Blood Count: CBC.*

1. ພາກສະເໜີ

ປັດຈຸບັນປະລິມານຂໍ້ມູນກ່ຽວກັບສຸຂະພາບນັບມື້ນັບເພີ່ມຂຶ້ນເລື້ອຍໆ ດັ່ງນັ້ນຜູ້ຄົ້ນຄວ້າຈຶ່ງໄດ້ມີແນວຄິດຢາກສຶກສາຂໍ້ມູນການກວດສຸຂະພາບໂດຍຜ່ານການກວດເລືອດ (ໂລຫິດວິທະຍາ) ເຊິ່ງເປັນຂັ້ນຕອນໜຶ່ງໃນການວິນິໄສ ພະຍາດຈາກຄົນເຈັບເປັນຕົວຊ່ວຍໃນການລະບຸອາການ ແລະ ສາເຫດຂອງການເກີດພະຍາດຕ່າງໆ, ເປັນປະໂຫຍດ ຫຼາຍໃຫ້ແພດໝໍ ຫຼື ຜູ້ຊ່ຽວຊານໃນການປິ່ນປົວພະຍາດ ຂອງຜູ້ປ່ວຍ. ເພື່ອປະໂຫຍດ ແລະ ຊ່ວຍໃນການ ຄາດຄະເນ ພະຍາດໂດຍການນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ເປັນຂັ້ນຕອນການວິເຄາະຂໍ້ມູນທີ່ມີປະລິມານ ມະຫາສານ (Big Data) ນິຍົມໃຊ້ໃນການຜັດທະນາຂໍ້ມູນ ຂ່າວສານໃນຫຼາຍຂະແໜງການໃນນັ້ນຂະແໜງການແພດ ເພາະວ່າມັນເປັນຕົວຊ່ວຍໃນການຕັດສິນໃຈ ແລະ ຄາດ ການຜົນທີ່ຈະໄດ້ຮັບຈາກການຕັດສິນໃຈ, ເພີ່ມຄວາມໄວ ໃນການວິເຄາະຖານຂໍ້ມູນຂະໜາດໃຫຍ່.

ການຄາດຄະເນຜົນການເກີດເຫດການລ່ວງໜ້າໄດ້ນັ້ນ ຈະເປັນປະໂຫຍດຫຼາຍຕໍ່ຂະແໜງການທາງດ້ານເສດຖະກິດ-ສັງຄົມ, ດ້ານການລົງທຶນ, ດ້ານການແພດ ແລະ ດ້ານອື່ນໆອີກ. ໂດຍການນຳເອົາຂໍ້ມູນທີ່ມີຢູ່ແລ້ວມາເປັນຕົວຊ່ວຍໃຫ້ຜູ້ຊ່ຽວຊານວິເຄາະ ແລະ ວາງແຜນການລ່ວງໜ້າເພື່ອເພີ່ມປະສິດຕິພາບ ເຊິ່ງຂໍ້ມູນທາງດ້ານການກວດເລືອດນີ້ຈະໄດ້ນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ເຊິ່ງປະກອບດ້ວຍເຕັກນິກການຈຳແນກປະເພດຂໍ້ມູນ (Classification), ນຳໃຊ້ການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree), ນຳໃຊ້ເຄືອຂ່າຍປະສາດທຽມ (Neural Network), ນຳໃຊ້ການຈຳແນກແບບ Naive Bayes, ນຳໃຊ້ໂປຣແກຣມເຄື່ອງມື Orange Tool ເປັນເຄື່ອງມືໃນການຂຸດຄົ້ນຂໍ້ມູນ ແລະ ເອົາຂໍ້ມູນການກວດເລືອດ Complete Blood Count: CBC ເຊ່ິ ງມາເປັ ນຂໍ້ ມູນ Dataset ທີ່ໄດ້ມາຈາກໂຮງໝໍ 5 ເມສາ ນັບຕັ້ງແຕ່ປີ 2019-2020.

ໂຮງໝໍ 5 ເມສາ ປະກອບມີທັງໝົດ 17 ພະແນກຄື: ພະແນກລັງສີວິທະຍາ (ປິ່ນປົວຄົນເຈັບໂດຍການນຳໃຊ້ເຄື່ອງສ້ອງລັງສີ ແລະ ເຄື່ອງສ້ອງເອໂກ), ພະແນກບໍລິຫານຈັດຕັ້ງ-ສັງລວມ (ຄຸ້ມຄອງອຸປະກອນເຄື່ອງໃຊ້ພາຍໃນໂຮງໝໍ, ການເງິນ, ລິດຂຶນສິ່ງຄົນເຈັບ ແລະ ຊັບຊ້ອນພະນັກງານເລື່ອນຂັ້ນຊັ້ນຕຳແໜ່ງ), ພະແນກພາຍນອກ (ປິ່ນປົວຄົນເຈັບທີ່ມີອາການເຈັບເປັນທາງພາຍນອກຮ່າງກາຍ ແລະ ຜ່າຕັດຂະໜາດນ້ອຍ-ໃຫຍ່), ພະແນກປິ່ນປົວເດັກ (ປິ່ນປົວພະຍາດເດັກນ້ອຍໂດຍສະເພາະ), ພະແນກປິ່ນປົວພະນັກງານ (ປິ່ນປົວພະນັກງານຕຳຫຼວດ ຫຼື ຄອບຄົວ), ພະແນກພາຍໃນ (ປິ່ນປົວຄົນເຈັບທີ່ເປັນພະຍາດທາງໃນຮ່າງກາຍຕົວຢ່າງ ພະຍາດວັນນະໂລກ, ໄຂ້ຍຸ້ງເປັນຕົ້ນ), ພະແນກພະຍາບານ (ຄຸ້ມຄອງພະຍາບານທັງໝົດພາຍໃນໂຮງໝໍ), ພະແນກວາງຢາສະຫຼົບ-ຟື້ນຟູຊີບ (ເປັນພະແນກຜູ້ຊ່ວຍທ່ານໝໍໃນການວາງຢາສະຫຼົບເພື່ອຜ່າຕັດຕ່າງໆ ແລະ ເປັນບ່ອນພັກຟື້ນຄົນເຈັບຫຼັງຜ່າຕັດ), ພະແນກແຂ້ວ (ປິ່ນປົວພະຍາດກ່ຽວກັບແຂ້ວໂດຍສະເພາະ), ພະແນກກວດເຂດນອກ-ສຸກເສີນ (ປິ່ນປົວຄົນເຈັບຈາກອາການເບື້ອງຕົ້ນໂດຍມີໜ້າທີ່ສົ່ງໄປປິ່ນປົວແຕ່ລະແຜນກຕາມອາການຄົນເຈັບ), ພະແນກມໍລະສຸມ (ປິ່ນປົວຄົນເຈັບທີ່ມີອາການຮຸນແຮງສຽງຕໍ່ການເສຍຊີວິດ), ພະແນກວິເຄາະ (ມີໜ້າທີ່ນຳຜົນການກວດເລືອດ, ຍ່ຽວ ແລະ ອາຈົມມາວິ ເຄາະຫາພະຍາດ), ພະແນກອອກລູກ-ພະຍາດຍິງ-ແມ່ ແລະ ເດັກ (ເປັນພະແນກເກີດລູກ, ພັກຟື້ນຄົນເຈັບຫຼັງຈາກເກີດລູກ ແລະ ປິ່ນປົວພະຍາດຂອງຜູ້ຍິງ), ພະແນກການຢາ (ເປັນບ່ອນສະໜອງຢາໃຫ້ຄົນເຈັບ ແລະ ຄຸ້ມຄອງຢາຕ່າງໆພາຍໃນໂຮງໝໍ), ພະແນກບໍລິຫານການແພດ (ຄຸ້ມຄອງແພດໝໍ, ພະຍາບານ, ຊັກລີດອຸປະກອນນຸ່ງຖືຂອງໂຮງໝໍ) ພະແນກກາຍະບຳບັດ (ປິ່ນປົວຄົນເຈັບດ້ວຍເຄື່ອງນວດໄຟຟ້າ, ຝັງເຂັ້ມ ແລະ ບຳບັດ), ພະແນກປະກັນສັງຄົມ (ຄຸ້ມຄອງພະນັກງານຜູ້ທີ່ມີບັດປະກັນສັງຄົມ).

ດັ່ງນັ້ນ, ໃນນາມຜູ້ຄົ້ນຄວ້າຈຶ່ງເຫັນວ່າການເຮັດວິໄຈໂດຍ ນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ເຂົ້າມາຊ່ວຍໃນດ້ານການແພດເພື່ອເປັນປະໂຫຍດ ແລະ ຊ່ວຍໃນການຄາດເດົາພະຍາດໃຫ້ໄດ້ຮັບຜົນດີ ນັ້ນບໍ່ແມ່ນເລື່ອງງ່າຍດາຍ, ຍັງເປັນສິ່ງທີ່ທ້າທາຍທັງໜ້າສົນໃຈຫຼາຍໃນກໍລະນີສຶກສາຄັ້ງນີ້ມີຄວາມສຳຄັນເປັນຢ່າງຍິ່ງ ເພາະນອກຈາກຈະຊ່ວຍໃຫ້ແພດໝໍປະຍັດເວລາ ແລະ ປະຍັດແຮງງານແລ້ວຍັງຊ່ວຍໃຫ້ມີຄວາມຈະແຈ້ງໃນການໃນການຄາດເດົາພະຍາດໃຫ້ໄດ້ຮັບຜົນດີ.

ຫົວຂໍ້ “ການຄາດຄະເນການເປັນເນື້ອງອກໃນເລືອດ ແລະ ຂໍ້ມູນຜົນການກວດເລືອດນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນ” ລະບົບການຮັກສາສຸຂະພາບຂອງມະນຸດນັ້ນມີຈຳນວນມະຫາສານເປັນຂໍ້ມູນໃຫຍ່ ການຂຸດຄົ້ນຂໍ້ມູນແມ່ນຂະບວນການຄິດໄລ່ຂອງຮູບແບບການຮຽນຮູ້ໃນຊຸດສະຖິຕິ ເຊັ່ນ: ພະຍາດເລືອດ, ມີຂໍ້ມູນການທົດສອບຫຼາຍຢ່າງທີ່ສາມາດເກັບໄດ້ຈາກຄົນເຈັບຂອງພວກເຂົາ. ບັນດາການຊອກຄົ້ນຫາ ແລະ ການສ້າງຂໍ້ມູນຈາກສະຖິຕິທາງການແພດ ໂດຍການນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນນັ້ນ ມີຄວາມຈຳເປັນໃນຂໍ້ມູນທາງການແພດຫຼາຍ, ເຊິ່ງສາມາດຊ່ວຍໃນການກວດຫາພະຍາດຕ່າງໆໃນໄລຍະເລີ່ມຕົ້ນ ຫຼື ກ່ອນເກີດພະຍາດນັ້ນ ໃນບົດຄົ້ນຄ້ວານີ້, ພວກເຂົາໄດ້ ນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນ ເພື່ອແກ້ໄຂການເຊື່ອມຕໍ່ລະຫວ່າງຄວາມເປັນເອກະລັກຂອງການກວດເລືອດ. (ໂລກເລືອດຈາງ) ແລະ ການໄຄ່ບວມຂອງເລືອດ, ເພື່ອບອກລ່ວງ

ໜ້າກ່ຽວກັບພະຍາດດັ່ງກ່າວໃນໄລຍະເລີ່ມຕົ້ນ, ເຊິ່ງສາມາດ ຜັດທະນາຄວາມສາມາດໃນການປິ່ນປົວ. ຈຸດປະສົ່ງຂອງ ການທົດລອງຂອງພວກເຂົາແມ່ນເພື່ອຜະລິດແມ່ແບບ ແລະ ແອັບທີ່ສາມາດແຈ້ງໃຫ້ຄົນເຈັບຮູ້ຜົນກວດ ສຳລັບປີດນີ້ຈະ ເປັນວິທີການ ແລະ ຂັ້ນຕອນການເຮັດຂຸດຄົ້ນຂໍ້ ມູນ (Abinaya & Sumathi, 2018)

ຫົວຂໍ້ "ຄາດຄະເນຄວາມສ່ຽງຂອງພາວະແຊກຊ້ອນ ຂອງໂລກເບົາຫວານ ໂດຍການນຳໃຊ້ເຕັກນິກການຂຸດຄົ້ນ ຂໍ້ມູນ" ໂລກເບົາຫວານເປັນໂລກໜຶ່ງທີ່ມີຄວາມຮ້າຍແຮງ ຫຼາຍ ແລະ ເຮັດໃຫ້ຄວາມດັນເລືອດສູງເປັນອັນຕະລາຍທີ່ ສຸດສຳລັບຄົນເຈັບທີ່ເປັນໂລກເບົາຫວານ. ໂລກເບົາຫວານ ສ່ວນຫຼາຍແມ່ນ ເກີດມາຈາກຮ່າງກາຍທີ່ຜະລິດຮໍໂມນ ແລະ ອິນຊູລິນບໍ່ພຽງພໍທີ່ຈະເອົານ້ຳຕານໃນເລືອດໄປສູ່ສ່ວນ ຕ່າງໆຂອງຮ່າງກາຍ, ສິ່ງຜົນໃຫ້ເກີດການອຸດຕັນຂອງນ້ຳ ຕານໃນເສັ້ນເລືອດແດງ ສິ່ງນີ້ສາມາດເຮັດໃຫ້ເກີດພະຍາດ ແລະ ອາການຕ່າງໆ. ອາການແຊກຊ້ອນໃນອະໄວຍະວະ ຕ່າງໆຕໍ່ໄປນີ້ນັກຄົ້ນຄວ້າສົນໃຈຮຽນຮູ້ກ່ຽວກັບອາການ ແຊກຊ້ອນຂອງ Hypertension ທີ່ມັກຈະເກີດຂຶ້ນພ້ອມ ກັນ. ໃນຜູ້ປ່ວຍໂລກເບົາຫວານທີ່ມີຄວາມດັນເລືອດສູງ ພວກເຂົາໄດ້ຄົ້ນຄວ້າສຶກສາຂໍ້ມູນຂອງຜູ້ເປັນໂລກເບົາ ຫວານກັບຜົນການກວດສຸຂະພາບເປັນປະຈຳປີ ດ້ວຍການ ວັດແທກຄວາມດັນ ດ້ວຍເຕັກນິກການຈັດແບ່ງປະເພດຂໍ້ ມູນ (Classification) ຈາກຮູບແບບແຜນວາດຕົ້ນໄມ້ຕັດ ສິນໃຈ (Decision Tree) ໂດຍການຈັດແບ່ງ Class (k-NN) ໃນການວິເຄາະຜົນ ແລະ ຄາດເດົາຄວາມສ່ຽງຂອງພະ ຍາດຂອງຄົນເຈັບທີ່ມີຄວາມດັນເລືອດສູງ, ໂດຍວິທີການ ຮູບແບບແຜນວາດຕົ້ນໄມ້ຕັດສິນໃຈ (Decision Tree) ມີ ຄວາມຈະແຈ້ງຫຼາຍກວ່າການຈັດແບ່ງ Class (k-NN) ສູງ ເຖິງ 92.08% (Rungruedee & Chumwatana, 2017)

ຫົວຂໍ້ "ການປຽບທຽບເຕັກນິກການຈັດປະເພດ ແຕກຕ່າງກັນໂດຍໃຊ້ WEKA ສຳລັບຂໍ້ມູນຜົນການກວດ ເລືອດ" ຜູ້ຊ່ຽວຊານດ້ານການແພດຕ້ອງການວິທີການຄາດ ເດົາທີ່ໜ້າ ເຊື່ອຖືໃນການວິນິໄສຂໍ້ມູນກ່ຽວກັບຜົນການ ກວດເລືອດ. ເພື່ອຫຼຸດຜ່ອນຂໍ້ມູນຈຳນວນຫຼວງຫຼາຍກ່ຽວ ກັບຄົນເຈັບ ແລະ ສະພາບການປິ່ນປົວທາງການແພດຂອງ ພວກເຂົາ. ໂດຍທົ່ວໄປແລ້ວໃນການຂຸດຄົ້ນຂໍ້ມູນ (ບາງຄັ້ງ ເອີ້ນວ່າການຂຸດຂໍ້ມູນ ຫຼື ການຄົ້ນພົບຄວາມຮູ້) ເພີ່ນສຶກ ສາກ່ຽວກັບສູດການຄິດໄລ່ການຈັດປະເພດຕ່າງໆ. ທິດສະດີ ຕົ້ນຕຳຈຸດປະສົ່ງເພື່ອສະແດງໃຫ້ເຫັນ ເຕັກນິກການປຽບ ທຽບຂອງສູດການຄິດໄລ່ການຈັດປະເພດຮູບແບບທີ່ແຕກ ຕ່າງກັນ ໂດຍໃຊ້ Waikato Environment for Knowledge ເປັນເຄື່ອງມືໃນການວິເຄາະຜົນ ຫຼື ເອີ້ນໂດຍຫຍໍ້ວ່າ ໂປຣແກຣມ WEKA ແລະ ຄົ້ນພົບວ່າສູດໃດທີ່ເໝາະສົມທີ່ ສຸດ ສຳລັບຜູ້ໃຊ້ທີ່ກຳລັງເຮັດວຽກກ່ຽວກັບຂໍ້ມູນຜົນການ ກວດເລືອດ ໃນການນຳໃຊ້ຮູບແບບທີ່ສະເໜີໃຫ້ ທ່ານໝໍ ຄົນໃໝ່ ຫຼື ຄົນເຈັບສາມາດຄາດເດົາຂໍ້ມູນກ່ຽວກັບຜົນການ ກວດເລືອດໄດ້ຜ່ານ Mobile App ທີ່ສາມາດປິ່ງມະຕິ ແລະ ສະແດງຜົນກ່ຽວກັບຂໍ້ມູນການກວດເລືອດ. ສູດການຄິດໄລ່ ທີ່ດີທີ່ສຸດໂດຍອີງໃສ່ຂໍ້ມູນກ່ຽວກັບເມັດເລືອດແມ່ນການ ຈັດປະເພດແບບຕົ້ນໄມ້ J48 ໂດຍມີຄວາມຖືກຕ້ອງ 97.16% ແລະ ເວລາທັງໝົດທີ່ໃຊ້ໃນການກໍ່ສ້າງຕົວແບບ ແມ່ນຢູ່ 0.03 ວິນາທີ, ສ່ວນປະເພດແບບ Naïve Bayes ໂດຍມີຄວາມຖືກຕ້ອງ 29.71% (Md. Nurul Amin & Md. Ahsan Habib, 2015)

ຫົວຂໍ້ "ການປຽບທຽບປະສິດຕິພາບໃນການຄາດຄະ ເນໂອກາດການເປັນໂລກເບົາຫວານ" ການສຶກສາປຽບທຽບ ປະສິດທິພາບໂອກາດການເກີດຂອງ ພະຍາດເບົາຫວານສຳ ລັບໂຮງໝໍແຫ່ງໜຶ່ງ ໂດຍໃຊ້ວິທີການຈັດປະເພດ 6 ແບບ ຄື: ວິທີການຫາຄວາມໃກ້ຄຽງ (k-nearest neighbor), ຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision tree), ວິທີເຄືອຂ່າຍ ປະສາດທຽມ (Artificial neural network), ວິທີການ ແບບ Support vector machine (SVM), ວິທີການ ແບບໄບນາລີ (Binary logistic regression) ແລະ ວິທີ ການ Naïve Bayes ໃນການປຽບທຽບຜົນຂອງວິທີຈຳ ແນກກຸ່ມທັງໝົດທັງ 6 ວິທີຈະໃຊ້ຄ່າຄວາມຊັດເຈນສະເລ່ຍ Mean Absolute Error (MAE) ແລະ ຄ່າຄວາມຄາດ ເຄື່ອນກຳລັງສອງສະເລ່ຍ Mean Square Error (MSE) ຜົນການສຶກສາພົບວ່າວິທີການເຄືອຂ່າຍປະສາດທຽມ ມີຄ່າ ຄວາມຖືກຕ້ອງ ຄ່າຄວາມຊັດເຈນສະເລ່ຍ ແລະ ຄ່າຄວາມ ຄາດເຄື່ອນກຳລັງສອງສະເລ່ຍດີທີ່ສຸດຄື 95.94%, 0.0491% ແລະ 0.0396% ຕາມລຳດັບ (Saichon Sinsomboonthong, 2018).

ຈຸດປະສົງຂອງການຄົ້ນຄວ້ານີ້ແມ່ນ:

- ເພື່ອຄາດຄະເນໂອກາດເກີດພະຍາດຈາກຜົນການກວດເລືອດໃນຮູບແບບການຝຶກສອນ.
- ເພື່ອສຶກສາຂັ້ນຕອນຫຼັກການ ແລະ ວິທີການໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນ (Data mining).
- ເພື່ອສຶກສາ ແລະ ປຽບທຽບການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree), ເຄືອຂ່າຍປະສາດທຽມ (Neural Network) ແລະ ການຈຳແນກປະເພດແບບ Naive Bayes.
- ເພື່ອພັດທະນາຮູບແບບ (Model) ໄປນຳໃຊ້ເຂົ້າໃນການສ້າງໂປຣແກຣມ.

2. ອຸປະກອນ ແລະ ວິທີການ

ໃນການຄົ້ນຄ້ວາໃນຄັ້ງນີ້ແມ່ນກຳນົດເອົາຜົນຂໍ້ມູນການກວດເລືອດໃນຊ່ວງປີ 2019-2020 ຈຳນວນ 1543 ຄົນທີ່ຜ່ານການກວດເລືອດໂຕຈິງ, ນຳໃຊ້ເຕັກນິກການປະມວນຜົນ ແລະ ການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ນັ້ນໄດ້ນຳໃຊ້ 3 ຮູບແບບຄື: ການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree), ເຄືອຂ່າຍປະສາດທຽມ (Neural Network) ແລະ ການຈຳແນກປະເພດແບບ Naive Bayes. ສ່ວນໂປຣແກຣມ Orange ໃຊ້ເຮັດ Train Model ແລະ Test Model ເພື່ອຫາຄວາມຊັດເຈນຂອງ Model ແລ້ວເອົາຮູບແບບທີ່ດີທີ່ສຸດມາເປັນໂຕຕົ້ນແບບໃນການສ້າງໂປຣແກຣມຄຳນວນຜົນການກວດເລືອດ (ຄວາມສ່ຽງທີ່ຈະເກີດພະຍາດ).

ໂດຍຜູ້ວິໄຈໄດ້ກຳນົດເອົາບັນດາຂໍ້ມູນທີ່ຜ່ານການລົງສຶກສາໂຕຈິງຈາກຂໍ້ມູນທີ່ຜ່ານເຄື່ອງວິເຄາະເມັດເລືອດມາແລ້ວ ມາເກັບໄວ້ໃນໂປຣແກຣມເກັບກຳຂໍ້ມູນ ລະອຽດໂດຍຈະສະແດງໃຫ້ເຫັນໃນຕາຕະລາງທີ 1

ເພື່ອຈະສຶກສາ ແລະ ຫາຂອບເຂດຄວາມສ່ຽງນັ້ນ ໄດ້ກຳນົດເອົາຄ່າມາດຕະຖານທີ່ຈະເປັນຄວາມສ່ຽງໄວ້ເປັນແຕ່ລະກໍລະນີດັ່ງນີ້:

ກໍລະນີ 1: ເມັດເລືອດແດງ (RBC) ຕ້ອງຢູ່ໃນລະດັບ 4.2-5.5 10^6/ul ຖ້າໜ້ອຍກວ່າ ຫຼື ສູງກວ່າຄ່າມາດຕະຖານ ຈະຢູ່ໃນສະຖານະມີຄວາມສ່ຽງ.

ກໍລະນີ 2: ເຮໂມໂກບິນ (HGB) ຕ້ອງຢູ່ໃນລະດັບ 12–16g/dl ຖ້າໜ້ອຍກວ່າ ຫຼື ສູງກວ່າຄ່າມາດຕະຖານ ຈະຢູ່ໃນສະຖານະມີຄວາມສ່ຽງ.

ກໍລະນີ 3: ເຮມາໂຕຊິດ (HCT) ຕ້ອງຢູ່ໃນລະດັບ 37-47 % ຖ້າໜ້ອຍກວ່າ ຫຼື ສູງກວ່າຄ່າມາດຕະຖານ ຈະຢູ່ໃນສະຖານະມີຄວາມສ່ຽງ.

ກໍລະນີ 4: ເມັດເລືອດຂາວ (WBC) ຕ້ອງຢູ່ໃນລະດັບ 5,000 - 11,000 10^3/ul ຖ້າໜ້ອຍກວ່າ ຫຼື ສູງກວ່າຄ່າມາດຕະຖານ ຈະຢູ່ໃນສະຖານະມີຄວາມສ່ຽງ.

ກໍລະນີ 5: ປະລິມານເກັດເລືອດ (PLT) ຕ້ອງຢູ່ໃນລະດັບ 140-400 10^3/ul ຖ້າໜ້ອຍກວ່າ ຫຼື ສູງກວ່າຄ່າມາດຕະຖານ ຈະຢູ່ໃນສະຖານະມີຄວາມສ່ຽງ.

ຜູ້ຄົ້ນຄ້ວາບົດນີ້ໄດ້ເຮັດໄປຕາມທິດສະດີຂອງເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນໂດຍໄດ້ເລືອກເອົາຂໍ້ມູນທີ່ຈຳເປັນ ແລະ ເປັນປະໂຫຍດຕັດເອົາຂໍ້ມູນທີ່ບໍ່ຄົບ ແລະ ຂໍ້ມູນທີ່ບໍ່ເປັນປະໂຫຍດອອກ ເພື່ອສະດວກໃນການໄປນຳໃຊ້ ເຊິ່ງຫຼັກໆ ແລ້ວເຮົາສາມາດຈຳແນກຂໍ້ມູນອອກເປັນ 3 ໝວດໃຫຍ່ໄດ້ແກ່: ເມັດເລືອດແດງ, ເມັດເລືອດຂາວ ແລະ ເກັດເລືອດ.

ຂັ້ນຕອນທີ່ຈະເຮັດ (ສຳລັບບົດຄົ້ນຄ້ວານີ້) ແມ່ນເອົາຂໍ້ມູນມາແລ້ວມາຜ່ານຂະບວນການສະກັດເອົາຂໍ້ມູນແບບ (Data Mining KDD) ໂດຍຜ່ານ 5 ຂັ້ນຕອນຫຼັກໆ ດັ່ງນີ້:

ຂັ້ນຕອນທີ 1: ໄດ້ເກັບກຳຂໍ້ມູນຜົນການກວດເລືອດ (Database Hematology) ໃນຫ້ອງວິເຄາະ.

ຂັ້ນຕອນທີ 2: ຈາກນັ້ນເອົາຂໍ້ມູນ ມາເຂົ້າຂະບວນການເຮັດຕາມທິດສະດີ ແລະ ຫຼັກການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ວັດແທກຂໍ້ມູນກ່ຽວກັບຕົວປ່ຽນຕ່າງໆທີ່ສົນໃຈເຂົ້າໃນລະບົບ, ເພື່ອໃຊ້ຂໍ້ມູນດັ່ງກ່າວເປັນການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ແລະ ຊ່ວຍໃຫ້ເຮົາສາມາດຕອບ-ຄຳຖາມທີ່ກຳລັງຄົ້ນຄ້ວາ ແລະ ກຳນົດແນວຄວາມຄິດຂອງການທົດສອບ ແລະ ປະເມີນຜົນ.

Data Cleaning: ຂັ້ນຕອນທີ່ຕ້ອງໄດ້ສະກັດເອົາຂໍ້ມູນທີ່ບໍ່ກ່ຽວຂ້ອງ ຫຼື ຂໍ້ມູນທີ່ບໍ່ປະໂຫຍດອອກ. (ຂໍ້ມູນທີ່ໄດ້ຕັດອອກໄປປະກອບມີຂໍ້ມູນລາຄາໃນການກວດ, ຂໍ້ມູນການກວດອາຈົມ (KOPA), ຂໍ້ມູນກວດນ້ຳຍ່ຽວ ແລະ ຂໍ້ມູນທີ່ບໍ່ຖືກຕ້ອງ ຫຼື ບໍ່ຄົບ).

Data Integration: ຂັ້ນຕອນການລວມເອົາຂໍ້ມູນຫຼາຍແຫຼ່ງເຂົ້າໃຫ້ເປັນຂໍ້ມູນຊຸດດຽວ (ໄດ້ນຳເອົາມູນໃນແຕ່ລະບ່ອນທີ່ມີການຈັດເກັບຜົນການກວດເລືອດຄື: ຫ້ອງວິເຄາະເລືອດ).

Data Selection: ຂັ້ນຕອນການດຶງເອົາຂໍ້ມູນສຳລັບການວິເຄາະຈາກແຫຼ່ງທີ່ບັນທຶກໄວ້ ເລືອກຂໍ້ມູນທີ່ ມີຄວາມສຳພັນກັນ ແລະ ເປັນປະໂຫຍດຕໍ່ການນຳໃຊ້.

Data Transformation: ຂັ້ນຕອນການປ່ຽນຂໍ້ມູນໃຫ້ເໝາະສົມກັບການນຳໃຊ້ໂດຍຜູ້ວິໄຈໄດ້ກຳນົດເອົາຂໍ້ມູນດັ່ງນີ້ ຂໍ້ມູນຜົນການກວດເລືອດປະກອບມີ: ເພດ (Sex), ເມັດເລືອດແດງ (Red Blood Cell Count: RBC), ເຮໂມໂກຼບິນ (Hemoglobin: HGB), ເຮມາໂຕຊິດ (Hematocrit: HCT), ເມັດເລືອດຂາວ (White Blood Cell Count: WBC), ເກັດເລືອດ (Platelets: PLT), ເຊິ່ງຜູ້ວິໄຈຈະນຳເອົາມູນໄປຈັດເກັບໃນຖານຂໍ້ມູນ.

ຂັ້ນຕອນທີ 3: ນຳຂໍ້ມູນທີ່ຜ່ານການເຮັດຊຸດຄົ້ນຂໍ້ມູນ (Data Mining) ມາແລ້ວຈັດເກັບໃນຖານຂໍ້ມູນເຊິ່ງຜູ້ຄົ້ນຄ້ວາໄດ້ເລືອກໃຊ້ໂປຣແກຣມ SQL ມາເປັນຕົວຈັດເກັບຂໍ້ມູນ.

ຂັ້ນຕອນທີ 4: ການນຳໃຊ້ຂໍ້ມູນໃນຄັ້ງນີ້ ຈະເປັນຮູບແບບການຮຽນຮູ້ແບບມີການສອນ (Supervised Learning) ແລະ ເລືອກນຳໃຊ້ Algorithm Of Data Mining ໃນການຄາດຄະເນຜົນທີ່ໄດ້ຮັບໂດຍເລືອກ 3 ແບບ Algorithm ຄື: ການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree), ເຄືອຂ່າຍປະສາດທຽມ (Neural Network) ແລະ ການຈຳແນກປະເພດແບບ Naive Bayes.

ຂັ້ນຕອນທີ 5: ສຸດທ້າຍແມ່ນນຳຂໍ້ມູນທັງໝົດມາເຂົ້າໂປຼແກຼມ Orange ຈະເປັນເຄື່ອງມືໃນການຫາຜົນຄັ້ງນີ້ ເຊິ່ງແມ່ນໄດ້ນຳເອົາຂໍ້ມູນທັງ 3 ຮູບແບບການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree), ເຄືອຂ່າຍປະສາດທຽມ (Neural Network) ແລະ ການຈຳແນກປະເພດແບບ Naive Bayes ນັ້ນມາເຂົ້າໂປຼແກຼມ Orange ເພື່ອມາຊ່ວຍທົດສອບ ແລະ ປະເມີນຜົນເພື່ອຫາຄວາມຈະແຈ້ງໃນການໃນການຄາດຄະເນຜົນທີ່ຈະຮັບ ແລະ ຫາປະສິດຕິພາບຂອງແຕ່ລະຮູບແບບ, ເພື່ອໄດ້ຮູບແບບທີ່ດີສຸດໄປເປັນໂຕຕົ້ນແບບໃນການພັດທະນາໂປຼແກຼມ, ດັ່ງສະແດງໃນຮູບທີ່ 1 ຢູ່ທ້າຍບົດ.

3. ຜົນໄດ້ຮັບ

ຂໍ້ມູນທັງໝົດ 1543 ຄົນ ຢູ່ໃນປີ 2019-2020 ເມື່ອຜ່ານການຂຸດຄົ້ນຂໍ້ມູນ (Data Mining) ໄດ້ຜົນເປັນຂໍ້ມູນຂອງເພດຊາຍ 1327 ທຽບເທົ່າກັບ (86%) ເພດຍິງ 216 ທຽບເທົ່າກັບ (14%) ສະເລ່ຍອາຍຸເທົ່າກັບ 47.4±6.8 ຜົນຄ່າຂອງການກວດເລືອດ CBC ມີຄ່າສະເລ່ຍຢູ່ໃນເກນປົກກະຕິ RBC, HGB, HCT, WBC, PTL ມີຄ່າສະເລ່ຍເທົ່າກັບ (51.98%) ແລະ ມີຄ່າ RBC, HGB, HCT, WBC, PTL ສະເລ່ຍຢູ່ໃນຄວາມສ່ຽງ (48.02%).

3.1 ຜົນຂໍ້ມູນທີ່ໄດ້ຈາກການຝຶກສອນ (Training Score)

▪ ການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree): ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Non-Risk ເທົ່າ 567 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 5, ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Risk ເທົ່າ 479 ແລະ ບໍ່ຖືກຕ້ອງເທົ່າ 30. (ຂໍ້ມູນ n= 1081, Non-Risk = 572, Risk = 509), ດັ່ງສະແດງໃນຕາຕະລາງທີ 3.

▪ ເຄືອຂ່າຍປະສາດທຽມ (Neural Network): ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Non-Risk ເທົ່າ 518 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 54, ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Risk ເທົ່າ 437 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 72. (ຂໍ້ມູນ n=1081, Non -Risk=572, Risk=509), ດັ່ງສະແດງໃນຕາຕະລາງທີ 4.

▪ ເນອິບ ເບ (Naive Bayes): ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Non-Risk ເທົ່າ 478 ແລະ ບໍ່ຖືກຕ້ອງເທົ່າ 94, ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Risk ເທົ່າ 393 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 116. (ຂໍ້ມູນ n=1081, Non-Risk = 572, Risk = 509), ດັ່ງສະແດງໃນຕາຕະລາງທີ 5.

3.2 ຜົນຂໍ້ມູນທີ່ໄດ້ຈາກການທົດສອບ (Testing Score)

▪ ການຕັດສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree): ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງ

ຂອງ Non-Risk ເທົ່າ 229 ແລະ ບໍ່ຖືກຕ້ອງ 0, ຜົນຂອງ ການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Risk ເທົ່າ 222 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 11. (ຂໍ້ມູນ n=462, Non-Risk = 229, Risk = 233), ດັ່ງສະແດງໃນຕາຕະລາງທີ 6.

▪ ເຄືອຂ່າຍປະສາດທຽມ (Neural Network): ຜົນຂອງການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Non-Risk ເທົ່າ 220 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 9, ຜົນຂອງ ການຄາດຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Risk ເທົ່າ 206 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 27. (ຂໍ້ມູນ n=462, Non-Risk = 229, Risk = 233), ດັ່ງສະແດງໃນຕາຕະລາງທີ 7.

▪ ເນອິບ ເບ (Naive Bayes): ຜົນຂອງການຄາດ ຄະເນຂໍ້ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Non-Risk ເທົ່າ 204 ແລະ ບໍ່ຖືກຕ້ອງ ເທົ່າ 25, ຜົນຂອງການຄາດຄະເນຂໍ້ ມູນ (Predicted) ທີ່ຖືກຕ້ອງຂອງ Risk ເທົ່າ 181 ແລະ ບໍ່ ຖືກຕ້ອງ ເທົ່າ 52. (ຂໍ້ມູນ n=462, Non-Risk = 229, Risk = 233) , ດັ່ງສະແດງໃນຕາຕະລາງທີ 8.

4. ວິພາກຜົນ

ໃນການສຶກສາວິໄຈຄັ້ງນີ້ເຫັນວ່າ ການຂຸດຄົ້ນຂໍ້ມູນ ນັ້ນມີຜົນດີຫຼາຍ ໃນດ້ານການສະກັດເອົາຄວາມຮູ້ໃນການ ວິໄຈຜົນໃນຄັ້ງນີ້ ຜົນການສຶກສາພົບວ່າ ມີອັດຕາຄວາມ ສ່ຽງສູງ ແລະ ປົກກະຕິບໍ່ມີຄວາມສ່ຽງ ໃນການສຶກສາຄັ້ງນີ້ ພົບວ່າມີຄວາມສອດຄ່ອງກັບການວິໄຈຂອງ THIPPALA (2017) ທີ່ສຶກສາຄົ້ນຄ້ວາກ່ຽວກັບເມັດເລືອດ ພົບວ່າເມັດ ເລືອດຕ່ຳ ແລະ ສູງເຮັດໃຫ້ມີຜົນຕໍ່ຄວາມສ່ຽງ ກໍ່ໃຫ້ເກີດ ເປັນໂລກຕຸ້ຍ ໂດຍເພີ່ມທຽບການຜິດປົກ ກະຕິຂອງເມັດ ເລືອດກັບຄ່າ BMI (BMI ແມ່ນການຄຳນວນອາຍຸ, ລວງ ສູງ ແລະ ນ້ຳໜັກ), ຈຸດເດັ່ນຂອງບົດວິໄຈນີ້ແມ່ນໄດ້ໂປຣ ແກຣມການວິເຄາະຫາຄວາມສ່ຽງຈາກເມັດເລືອດ.

ຜົນການສຶກສາຄັ້ງນີ້ ໄດ້ນຳໃຊ້ຮູບແບບການຕັດສິນ ໃຈແບບຕົ້ນໄມ້, ເຄືອຂ່າຍປະສາດທຽມ, ເນອິບເບ ເຊິ່ງພົບ ວ່າການເຮັດແບບຕົ້ນໄມ້ໃນການຫາຄວາມສ່ຽງໃຫ້ຜົນຄ່າ ຄວາມຊັດເຈນກວ່າໝູ່ ເຊິ່ງພົບວ່າສອດຄ່ອງກັບການວິໄຈ ຂອງ Saichon Sinsomboonthong (2018) ໂດຍບົດ ຄົ້ນຄ້ວາເພີ່ນແມ່ນນຳເອົາຂໍ້ມູນເລືອດຂອງກຸ່ມຄົນທີ່ເປັນ ໂລກໄຕ ມາທົດລອງຫາປະສິດຕິພາບຫາຄວາມຊັດເຈນ ຂອງຄວາມສ່ຽງທີ່ຈະເປັນໂລກໄຕເຊິ່ງໄດ້ນຳເອົາຫຼາຍວິທີ ລວມເຖິງວິທີຂອງການສິນໃຈແບບຕົ້ນໄມ້ ແລະ ຜົນມີ ຄວາມຊັດເຈນເຖິງ 100% ສ່ວນບົດຜູ້ວິໄຈໄດ້ຄ່າແບບ ການຕັດສິນໃຈແບບຕົ້ນໄມ້ 97.1%

ພ້ອມນີ້ຜູ້ວິໄຈໃນຄັ້ງນີ້ບໍ່ໄດ້ຈັດວ່າ ລະດັບໃດຈຶ່ງສ່ຽງ ເປັນພະຍາດຫຍັງ ພຽງແຕ່ຫາຄວາມສ່ຽງຂອງຜົນການກວດ ເລືອດທີ່ ມີຄ່າສູງ-ຕ່ຳ ແລະ ປົກກະຕິ (Thima 2010) (Sirin , Nurunart ແລະ Suntree 2021) ເປັນການວິໄຈ ຫາກຸ່ມເມັດເລືອດທີ່ເຈາະຈົງເອົາພຽງແຕ່ຄ່າຂອງ RBC, HGB, HCT, WBC ແລະ PLT ເທົ່ານັ້ນເພື່ອສຶກສາຫາ ຄວາມສ່ຽງ ໃນການສຶກສາຄັ້ງນີ້ນອກຈາກຈະຮູ້ຄ່າມາດຕະ ຖານຄ່າຄວາມສ່ຽງແລ້ວ ແຕ່ກໍຍັງຂາດຫຼາຍແຫຼ່ງທີ່ມາສົມ ທົບກໍ່ໃຫ້ເກີດພະຍາດ ຖ້າຢາກໃຫ້ຮູ້ຊັດເຈນແທ້ວ່າຜົນຖືກ ຕ້ອງທີ່ສຸດແມ່ນຕ້ອງໄດ້ອາໄສ ຫຼື ອີງໃສ່ໃຫ້ແພດໝໍຜູ້ຊຳ ນານໃນການວິເຄາະຜົນ ເນື່ອງຈາກຄົນປ່ວຍຜູ້ໜຶ່ງຕ້ອງມີ ຫຼາຍສາເຫດໃນການເກີດພະຍາດຕ້ອງໄດ້ຖາມອາການ ເບື້ອງຕົ້ນ, ໂລກປະຈຳຕົວຂອງຜູ້ປ່ວຍ, ການກິນຢູ່ ແລະ ລວມທັງການກິນຢາເປັນຕົ້ນ.

5. ສະຫຼຸບ

ການສຶກສາໃນຄັ້ງນີ້ ໄດ້ເຮັດການວິເຄາະຂໍ້ມູນ ຜົນ ຄ່າສົມບູນຂອງການກວດເລືອດ CBC ສາມາດໄດ້ຂໍ້ມູນທີ່ ຜ່ານການເຮັດ Data Minnig ມາແລ້ວເປັນ Data Set ທີ່ ປະກອບໄປດ້ວຍຂໍ້ມູນເກນເລືອດໃນລະດັບປົກກະຕິ 801 ຄົນ ແລະ ຂໍ້ມູນຜົນການກວດເລືອດໃນເກນທີ່ມີຄວາມສ່ຽງ 742 ຄົນ ລວມທັງໝົດ 1543 ຄົນ ໂດຍຜ່ານການທົດສອບ ມູນທັງ 3 ຮູບແບບ ໄດ້ສະແດງໃຫ້ເຫັນໄດ້ວ່າຂໍ້ມູນ 1543 ຄົນ ນີ້ມີຄ່າຄວາມສ່ຽງ ແລະ ບໍ່ສ່ຽງ ໃນຮູບແບບການຕັດ ສິນໃຈແບບຕົ້ນໄມ້ (Decision Tree) ມີຄວາມຊັດເຈນສູງ , ຮູບແບບເຄືອຂ່າຍປະສາດທຽມ (Neural Network) ມີ ຄວາຊັດເຈນສູງ ແລະ ຮູບແບບ ເນອິບ ເບ (Naive Bayes) ມີຄວາມຊັດເຈນສູງ, ດັ່ງສະແດງໃນຕາຕະລາງທີ 9. ສະນັ້ນ ສະຫຼຸບໄດ້ວ່າ ຮູບແບບການຕັດສິນໃຈແບບຕົ້ນໄມ້ມີຄວາມ ຊັດເຈນສູງກວ່າ.

6. ຂໍ້ຂັດແຍ່ງ

ຂ້າພະເຈົ້າໃນນາມຜູ້ຄົ້ນຄວ້າວິທະຍາສາດ ຂໍປະຕິ ຍານຕົນວ່າ ຂໍ້ມູນທັງໝົດທີ່ມີໃນບົດຄວາມວິຊາການດັ່ງ

ກ່າວນີ້ ແມ່ນບໍ່ມີຂໍ້ຂັດແຍ່ງທາງຜົນປະໂຫຍດກັບພາກສ່ວນໃດ ແລະ ບໍ່ໄດ້ເອື້ອປະໂຫຍດໃຫ້ກັບພາກສ່ວນໃດພາກສ່ວນໜຶ່ງ, ກໍລະນີມີການລະເມີດ ໃນຮູບການໃດໜຶ່ງ ຂ້າພະເຈົ້າມີຄວາມຍິນດີ ທີ່ຈະຮັບຜິດຊອບແຕ່ຜຽງຜູ້ດຽວ.

7. ເອກະສານອ້າງອີງ

Abinaya, & Sumathi. (2018). International Journal of Computer Science Trends and Technology (IJCST). Predicting Blood Tumor and Hematology Using Data Mining Techniques, 20-22.

Demšar, J., & Zupan, B. (2012). Orange. Data Mining Fruitful And Fun, 45-46. Retrieved from https://scholar.google.com/scholar?q=Orange:+Data+Mining&hl=th&as_sdt=0&as_vis=1&oi=scholart

Hans-Petter, H. (2017). Structure Query Language. norway. Retrieved from The SQL language: https://www.halvorsen.blog/

Jiawei, H., Micheline, K., & Jian, P. (2012). Data Mining Concepts and Techniques Third Edition. United States of America. Retrieved from http://myweb.sabanciuniv.edu › files › 2016/02

Md. Nurul Amin, & Md. Ahsan Habib. (2015). Comparison of Different Classification Techniques Using WEKA for Hematological Data. Mawlana Bhashani Science and Technology.

Orawan, L. (2016). Classification of Facial Expressions of Cartoons using Histogram of Oriented Gradient. 41-42. Retrieved from http://ethesisarchive.library.tu.ac.th › thesis

Pacharawongsakda, E. (2015). Cross Validation. Bangkok: Co-Founder and Lead Data Scientist at Cube Analytics Consulting. Retrieved from K-fold Cross Validation: https://th.linkedin.com/in/eakasit-pacharawongsakda-ph-d-475a8452

Pagon Gatchalee. (2019). Confusion Matrix. Chaing Mai University: Computer Software and Theory. Retrieved from https://medium.com/@pagongatchalee

Pornpon, T. (2009). classification using feature reduction and support vector machine, 17-18. Retrieved from https://kb.psu.ac.th › psukb › bitstream

Richard, R., & Michael, G. (2003). Data Mining: A Tutorial Based Primer 1st Edition. United States of America.

Rungruedee, B., & Chumwatana, T. (2017). Prediction on risk for complications of hypertension in diabetes patient using data mining technique. College of Information and Communication Technology, Rangsit University, Thailand.

Saichon Sinsomboonthong. (2018). An Efficiency Comparison of Diabetes Prediction. Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Road, Ladkrabang, Bangkok, Thailand.

Sirin , F., Nurunart, Y., & Suntree, K. (2021). Annual Health Checkup . Understanding And Interpreting Of Preliminary Laboratory Results For Annual Health Checkup, 124-125.

Suphamon , C. (2018). Data Mining. Data Mining Techniques for Nursing Data Analysis, 84-85.

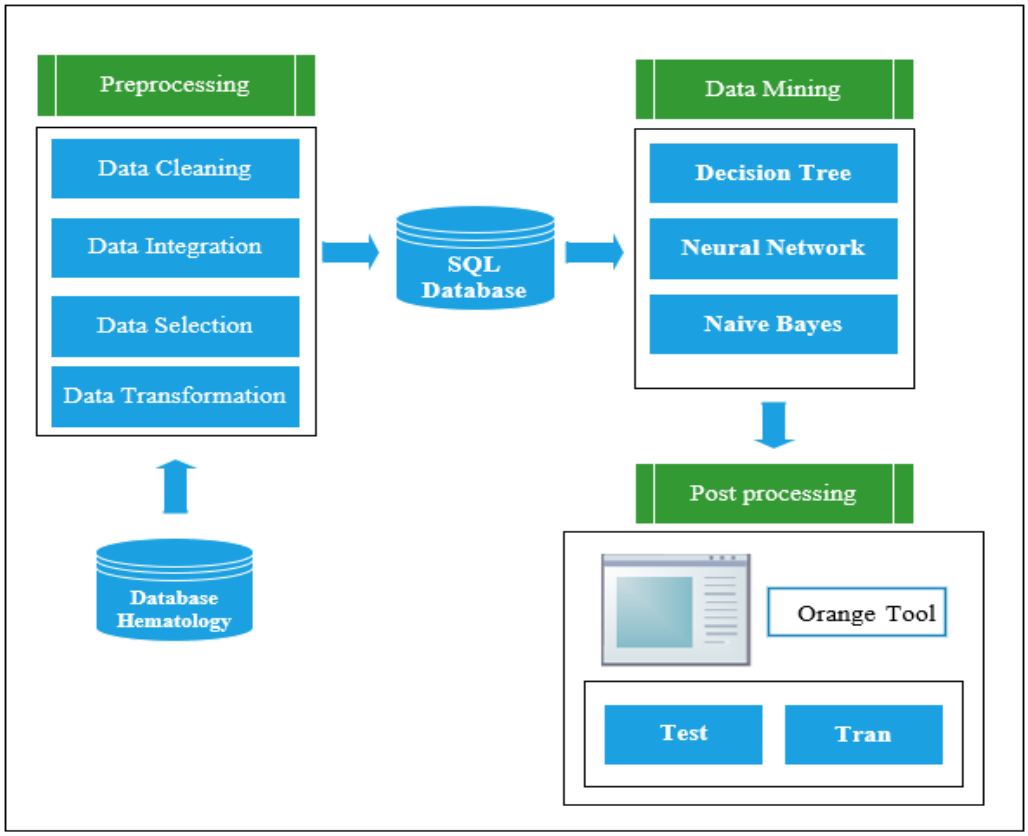
Thima, S. (2010). CBC. Complete Blood Count. Retrieved from Complete Blood Count: std.kku.ac.th/4831500146/narok/PathoImmunopathology%20I.doc , www.siamhealth.net/public_html/Health/Lab_interprete/cbc.htm ,

THIPPALA, M. (2017). Associaion Between Body Mass Index And Blood Cell. Faculity Of Allied Health Sciences, Thammasat University, Thailand.

Udeme, U. (2018, February 13). Basic Overview of Convolutional Neural Network (CNN). Retrieved from Activation Function: https://medium.com/dataseries/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17

Ukrit , P. (2003). The project uses data mining techiques to predict the flow of water. University of Technology King Mongkut Thonburi.

ຮູບທີ 1: ແຜນພາບລວມແບບການສຶກສາ ແລະ ອະທິບາຍຜົນ.



ຕາຕະລາງທີ 1: ຂໍ້ມູນ ແລະ ລາຍລະອຽດທີ່ຈະນຳມາຄົ້ນຄ້ວາໃນບົດນີ້

ລ/ດ	ຂໍ້ມູນ	ຄວາມໝາຍ	ລະດັບປົກກະຕິ	ຫົວໜ່ວຍ	ປະເພດ
1	Name	ຊື່	-	-	Text
2	Age	ອາຍຸ	-	-	Number
3	Sex	ເພດ	-	0,1	Number
4	RBC	ເມັດເລືອດແດງ	4.2 - 5.5	10^6/uL	Number
5	HGB	ເຮໂມໂກຼບິນ	12 – 16	g/dL	Number
6	HCT	ເຮມາໂຕຊິດ	37 – 47	%	Number
7	WBC	ເມັດເລືອດຂາວ	5,000 - 11,000	10^3/uL	Number
8	PLT	ປະລິມານເກັດເລືອດ	140-400	10^3/uL	Number
9	Level	ຄວາມສ່ຽງ	No-Risk, Risk	-	Text

ຕາຕະລາງທີ 2: ສະແດງຜົນຂໍ້ມູນ ແລະ ຜົນຂອງການກວດເລືອດ (CBC)

ຂໍ້ມູນທົ່ວໄປ	Min.	Max.	ຈຳນວນຂໍ້ມູນ (%)
ເພດ			1 = ຊາຍ, 0 = ຍິງ
ຊາຍ			1327 (86%)
ຍິງ			216 (14%)
ອາຍຸ (ປີ) mean±SD	29	70	47.4±6.8
RBC: Red blood cell mean±SD	3.55	8.1	5.17±0.671

HGB: Hemoglobin mean±SD	7.9	18.6	13.625±1.487
HCT: Hematocrit mean±SD	25	58	42.039.4±4.099
WBC: White blood cell mean±SD	2900	74000	6984.323±3021.995
PLT: Platelet count mean±SD	95	615	249.756±60.529
Level			
Non-Risk			801 (51.98%)
Risk			742 (48.02%)
(n = 1543)			

ຕາຕະລາງທີ 3: ຜົນຂອງ Confusion Matrix Test and Score ການຕັດສິນໃຈແບບຕົ້ນໄມ້

	Predicted	
Acutual	Non-Risk	Risk
Non-Risk	567	5
Risk	30	479

ຕາຕະລາງທີ 4: ຜົນຂອງ Confusion Matrix Test and Score ເຄືອຂ່າຍປະສາດທຽມ

	Predicted	
Acutual	Non-Risk	Risk
Non-Risk	518	54
Risk	72	437

ຕາຕະລາງທີ 5: ຜົນຂອງ Confusion Matrix Test and Score Naive Bayes

	Predicted	
Acutual	Non-Risk	Risk
Non-Risk	478	94
Risk	116	393

ຕາຕະລາງທີ 6: ຜົນຂອງ Confusion Matrix Predictions ການຕັດສິນໃຈແບບຕົ້ນໄມ້

	Predicted	
Acutual	Non-Risk	Risk
Non-Risk	229	0
Risk	11	222

ຕາຕະລາງທີ 7: ຜົນຂອງ Confusion Matrix Predictions ເຄືອຂ່າຍປະສາດທຽມ

	Predicted	
Acutual	Non-Risk	Risk
Non-Risk	220	9
Risk	27	206

ຕາຕະລາງທີ 8: ຜົນຂອງ Confusion Matrix Predictions Naive Bayes

	Predicted	
Acutual	Non-Risk	Risk
Non-Risk	204	25
Risk	52	181

ຕາຕະລາງທີ 9: ສະຫຼຸບຜົນລວມຂອງການຄົ້ນຄ້ວາ

	Accuracy%	F1%	Precision%	Recall%
Decision Tree	97.1	97.1	97.2	97.1
Neural Network	90.4	90.4	90.4	90.4
Naive Bayes	88.2	81.3	81.4	81.3