Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

Classification of Personality Traits based on Data Fusion using Artificial Intelligence

A Thesis Submitted to the Council of the College of Computer Science and Mathematics University of Mosul as a Partial Fulfillment of Requirements for the Degree of Doctor of Philosophy in Computer Science

By

Ruba Talal Ibrahim Ahmed

Supervised by

Prof. Dr. Fawziya Mahmood Ramo

2023 A.D. **1444 A.H.**

ABSTRACT

Personality traits reflect a person's thoughts, feelings, and behavior as well as his culture and social environment. Personality traits are quantitative measures, so they are based on general concepts. For example, we say that this person is tense, open, or eager for his appointment, and so on. There are many test personality measures, the most famous of which is the Big Five factor, which classifies an individual's personality into five categories: agreeableness, conscientiousness, extraversion, neuroticism, and openness. This factor shows the personality dimensions of a group of people and the differences that come from their backgrounds or upbringings.

A signature is a reliable and effective predictor of personality based on its features established by graphology. It is easy to use, and most people have a single, and fixed signature. Also, signature analysis is a projection method similar to body language that sums up several aspects of a person's personality, like their social skills, how they think, and how they do their work. Therefore, it was used in this study to classify the personalities of individuals.

In this work, a multimodal intelligent system for classifying main personality traits, called Personality Traits Analysis Intelligent System (PTAIS) was built. The intelligent system consists of several steps. In the first step, samples of signature images of University of Mosul faculty members and their demographic information, as well as samples of their personal analysis questions were collected. In the second step, a personality analysis dictionary (RTFM) was created to categorize the dataset into five classes based on the Big Five factor. Then, a personality science expert examined each of these five groups. In the third step, the dataset is first processed, and then structural, statistical, and deep learning algorithms are used to extract the important features. The last step was to use four intelligence models to classify personality traits into five different groups.

The first model was called 3MPPT (Three Models Pretrained Personality Traits). This model used three separate pre-trained models (VGG16, ResNet50, and Inceptionv3) to classify personality traits after freezing top layers and adjusting bottom layers. The PCNN (Proposed Convolutional Neural Network) was the second model. It reconstructed a convolutional neural network by adjusting its layers, nodes, filters, and hyperparameters. These two models used regularization methods like dropout, learning rate and Adam optimization method. The third model (PXGBOOST) was responsible for classifying the fusion dataset, which is extracted by several algorithms and then selected the most important features

using the featurewiz library. Then, the features were categorized into groups using the parallel XGBOOST algorithm and four CPU cores to exploit the most computer resources. The latest model was called HFDL5 (Hybrid Fuzzy Deep Learning Five). This model has been used to modify deep learning by combining it with fuzzy learning and making fuzzy learning a layer within the deep learning layers. In this model, two different kinds of features were taken out, and the benefits of both were used to get rid of ambiguity and noise.

The results were analyzed and discussed using different performance and statistical methods. The results showed that the proposed models (HFDL5, PXGBOOST, and PCNN) did the best on all measures, especially the accuracy and Cohen's kappa measures compared to the pre-trained models, traditional methods, and the literature review. The hybrid model (HFDL5) achieved 99%, 98% in accuracy and Cohen's kappa measure respectively. The PXGBOOST model performed 98% of accuracy and 97% Cohen's kappa measure. While the PCNN model had accuracy of 93% and 75% Cohen's kappa measure. The PCNN did better than all the other models except the ResNet50 and Inceptionv3 models. The first model had 97% of accuracy and 88% of Cohen's kappa. The second model slightly exceeded the PCNN model as it had 95% of accuracy and 83% of Cohen's kappa score.

Finally, the results revealed that there is a good similarity and a relationship between signature analysis and the analytical question forms by 71%. The results also showed that the highest category of faculty members at the University of Mosul who signatures were collected is conscientious.

وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

تصنيف سمات الشخصية بالاعتماد على دمج البيانات باستخدام الذكاء الاصطناعي

اطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دكتوراه فلسفة في
علوم الحاسوب

من قبل

ربى طلال إبراهيم احمد

بإشراف
أ.د فوزية محمود رمو

١٤٤٤ هـ ٢٠٢٣ م

الخلاصة

تعكس سمات الشخصية أفكار ومشاعر وسلوك الشخص وكذلك ثقافته وبيئته الاجتماعية. سمات الشخصية هي مقاييس كمية لذا فهي تستند إلى مفاهيم عامة. على سبيل المثال، نقول إن هذا الشخص متوتر، او منفتح، أو حريص على موعده، وهكذا. هناك العديد من مقاييس الاختبار لقياس الشخصية، وأشهرها عامل الخمسة الكبار (Big Five Factor)، الذي يصنف شخصية الفرد إلى خمس فئات: التوافق، والضمير، والانبساط، والعصابية، والانفتاح. يوضح هذا العامل أبعاد شخصية مجموعة من الناس والاختلافات التي تأتي من خلفياتهم أو نشأتهم.

ويعد التوقيع مؤشر موثوق وفعال للشخصية بناءً على ميزاته التي أنشأها علم دراسة الخط. كما يعد سهل الاستخدام، ومعظم الناس لديهم توقيع واحد وثابت. ان تحليل التوقيع هو طريقة إسقاط تشبه لغة الجسد تلخص العديد من جوانب شخصية الفرد، مثل مهاراته الاجتماعية، وطريقة تفكيره، وكيف يؤدي عمله. لذلك تم استخدامه في هذه الدراسة لتصنيف شخصيات الأفراد.

في هذا العمل، تم بناء نظام ذكائي متعدد التقنيات لتصنيف سمات الشخصية، سمي PTAIS (النظام الذكائي لتحليل سمات الشخصية). يتكون النظام الذكائي من عدة مراحل. المرحلة الأولى تضمنت جمع عينة من صور تواقيع أعضاء هيئة التدريس في جامعة الموصل والمعلومات الديموغرافية الخاصة بهم، إضافة إلى جانب نماذج لأسئلة التحليل الشخصية. في المرحلة الثانية، تم إنشاء قاموس تحليل الشخصية (RTFM) لتصنيف قاعدة البيانات الى خمس فئات بناءً على عامل الخمسة الكبار والتي تم فحصها من قبل خبير في علم الشخصية. اما في المرحلة الثالثة، تمت معالجة قاعدة البيانات أولاً، ومن ثم استخدام الخوارزميات الهيكلية والإحصائية والتعلم العميق لاستخراج الميزات المهمة. في المرحلة الأخيرة تم تصنيف سمات الشخصية إلى خمس مجموعات مختلفة باستخدام أربعة نماذج ذكائية .

استخدم النموذج الأول3MPPT(Three Models Pretrained Personality Trait) ثلاثة نماذج منفصلة مُدربة مسبقًا (VGG16)و(ResNet50)و(Inceptionv3) لتصنيف سمات الشخصية بعد تجميد الطبقات العليا وإعادة ضبط الطبقات السفلى. وكان PCNN (Proposed Convolutional Neural Network) النموذج الثاني الذي تم فيه اعادة بناء هيكلية شبكة عصبية تلافيفيه عن طريق تعديل طبقاتها وعقدها ومرشحاتها ومعلماتها الفائقة. استخدم هذا النموذج طرق التنظيم مثل dropout والتحسين مثل Adam ومتوسط التدرج العشوائيSGD. اما النموذج الثالث

(PXGBOOST)مسؤول عن تصنيف Fusion Dataset، والتي تم استخلاصها من خلال عدة خوارزميات، ثم تحديد أهم الميزات باستخدام مكتبة Featurewiz. بعد ذلك، تم تصنيف الميزات باستخدام خوارزمية XGBOOSTالمتوازية وأربعة أنوية CPU لاستغلال معظم موارد الكمبيوتر. النموذج الاخير يسمى HFDL5 (Hybrid Fuzzy Deep Learning Five) الذي طور التعلم العميق من خلال دمجه مع التعلم الغامض وجعله طبقة داخل طبقات التعلم العميق. في هذا النموذج، تم استخدام نوعين مختلفين من الميزات، والاستفادة من كلا التمثيلين للتخلص من الغموض والضوضاء.

تم تحليل النتائج ومناقشتها باستخدام الأساليب الادائية والاحصائية المختلفة. أظهرت النتائج أن النماذج المقترحة (HFDL5 وPXGBOOST وPCNN) حققت أفضل أداء في جميع المقاييس، لا سيما الدقة ومقياس كوهين كابا، مقارنة بالنماذج المدربة مسبقًا والطرق التقليدية والدراسات السابقة. حيث حقق النموذج الهجين (HFDL5) دقة تصل إلى ٩٩٪، وكانت نتيجة كوهين كابا له ٩٨٪. اما النموذج (PXGBOOST)فقد حقق دقة ٩٨٪، ونتيجة كوهين كابا ٩٧٪. حقق نموذج PCNN دقة ٩٣٪ وكوهين كابا ٧٥٪. تم ملاحظة ان نموذج PCNN حقق أداءً أفضل من جميع الموديلات الأخرى باستثناء الطرق ResNet50 وInceptionv3. كانت دقة النموذج الأول ٩٧٪ ودرجة كوهين كابا له ٨٨٪. كذلك تجاوز النموذج الثاني نموذج PCNN قليلاً، حيث كانت تبلغ دقته ٩٥٪ ودرجة كوهين كابا ٨٣٪.

وأخيراً كشفت النتائج أن هناك تشابهاً جيداً وعلاقة بين تحليل التوقيع وصيغ الأسئلة التحليلية بنسبة ٧١٪. كما أظهرت النتائج أن أعلى فئة من أعضاء هيئة التدريس في جامعة الموصل هي فئة الضمير.