

Uniwersytet Przyrodniczy we Wrocławiu
Ogłasza konkurs otwarty na okres 48 miesięcy
na stanowisko doktoranta w Katedrze Genetyki, Pracownia Biostatystyki

Osoba na tym stanowisku zatrudniona będzie w ramach projektu PreludiumBis
„Wykorzystanie metody głębokiego uczenia w analizie sekwencji genomu zwierząt
hodowlanych” (nr 2019/35/O/NZ9/00237)
zgodnie z harmonogramem badań.

Kandydat na stanowisko asystenta powinien spełniać następujące kryteria:

1. posiadać stopień naukowy licencjata lub magistra w dziedzinie nauk biologicznych lub technicznych;
2. wykazywać zaawansowaną umiejętność programowania w językach R i Python;
3. posiadać umiejętność pracy w systemie operacyjnym Linux (również w linii komend, bez interfejsu graficznego);
4. posiadać umiejętność instalowania bioinformatycznych narzędzi programistycznych zarówno komercyjnych, jak i open source;
5. posiadać znajomość metody statystyki matematycznej;
6. wykazywać się podstawowym dorobkiem naukowym w zakresie bioinformatyki, biostatystyki lub biologii obliczeniowej;
7. posiadać czynną znajomość języka angielskiego pozwalającą na płynną komunikację werbalną i pisemną;
8. posiadać umiejętność pracy w zespole;
9. wykazywać zainteresowanie prowadzeniem badań naukowych i rozwijaniem własnego warsztatu naukowego, również poza “sztywno” określonymi godzinami pracy i zadaniami badawczymi realizowanymi w ramach projektu.

Dodatkowe aspekty, które będą brane pod uwagę w przypadku niemożności wyłonienia kandydata na podstawie powyższych kryteriów to:

1. znajomość pakietu Keras;
2. umiejętność programowania w językach Fortran lub C.

Wykaz wymaganych dokumentów:

1. życiorys;
2. autoreferat naukowy z uwzględnieniem planów rozwoju na przyszłość;
3. odpis dyplomu uzyskania stopnia licencjata lub magistra.

Dokumentację należy składać drogą e-mailową na adres:

[**joanna.szyda@upwr.edu.pl**](mailto:joanna.szyda@upwr.edu.pl)

Termin składania zgłoszeń upływa z dniem: 06.07.2020 o godzinie 23:59

Termin rozstrzygnięcia konkursu: 10.07.2020

Termin rozpoczęcia pracy do uzgodnienia, nie wcześniej niż od 01.10.2020

Informacje o przebiegu projektu

1. Kandydat będzie studentem Szkoły Doktorskiej Uniwersytetu Przyrodniczego we Wrocławiu i będzie zobowiązany do uczestnictwa w zajęciach dydaktycznych realizowanych w ramach Szkoły oraz to realizacji dydaktyki w wymiarze określonym przez statut Szkoły.
2. Kandydat otrzyma stypendium doktoranckie począwszy najwcześniej od dnia 1 października 2020 r. w wysokości 5 000 zł miesięcznie do miesiąca, w którym została przeprowadzona ocena śródkresowa w Szkole Doktorskiej oraz 6 000 zł miesięcznie od miesiąca, w którym została przeprowadzona ocena (w przypadku oceny pozytywnej).
3. W czasie trwania studiów doktoranckich Kandydat jest zobowiązany do odbycia stażu zagranicznego, trwającego od 3 do 6 miesięcy w wybranym przez Kierownika Projektu zagranicznym ośrodku naukowym.
4. Skrócony opis projektu w j. angielskim:

The project is going to use DL for 4 different aspects of data analysis, all related to whole genome DNA sequence data as genomic input. For each data set the development of the final DL algorithm will involve: the selection of the optimal number of layers, the structure of each layer and the optimisation algorithm. The final parameter estimates will be obtained based on a K-fold cross-validation across the training data, eventually with the implementation of a bootstrap algorithm. The training data sets are already available prior to project initialisation, while the corresponding, independent test data sets are going to be acquired within the frame of the project. The implementation will take place on a Linux computing server with the Python's Keras library based on TensorFlow algorithms.

The 1st problem involves the classification of single nucleotide polymorphisms (SNP) from the whole genome sequence into true-positive and false-positive calls. The training data set consists of four bulls with SNP genotypes identified based on an oligonucleotide microarray (GeneSeek ® Genomic ProfilerTM HD) and based on the whole genome DNA sequence (WGS). False-positive calls are represented by SNPs with genotypes discordantly called by both methods. The predictor will involve sequencing quality and alignment statistics at the SNP site as well as a DNA sequence down- and up-stream of the SNP. The aim is to estimate the impact of the available sequence information on the probability of calling an incorrect (i.e. a false-positive) SNP. The 2nd problem involves the classification of cows into mastitis-resistant and mastitis-prone, using SNP and copy number variant (CNV) genotypes identified based on WGS of 32 Polish Holstein-Friesian cows with known clinical mastitis diagnosis. The quality of disease status prediction for the training data obtained by using different sub-sets of SNPs and CNVs will be compared in order to identify an optimal set resulting in high prediction accuracy. The next data set, comprises SNPs and indels identified from WGS represents data from 1000 Bull Genomes Project (run7) and consists of 3103 individuals representing Angus, Brahman, Brown Swiss, Charolais, Danish Red, Fleckvieh, Gelbvieh, Hereford, Holstein, Jersey, Limousin, Montbeliarde, Normande, and Simmental breeds. It will be used for problem 3 comprising assignment of bulls into breeds - a multilevel classification problem involving SNPs and indels as classifier variables. The same data will also be utilised for problem 4 for the context of the application of Recurrent Neural Networks for imputation of SNPs from the genotype resolution corresponding to the 50K and 777K BovineHD Genotyping BeadChips (Illumina) to the SNP resolution corresponding to WGS.