

Aufgabenstellung

Definitive Anmeldung zum Wirtschaftsprojekt/zur Bachelorarbeit

1. Starttermin:

spätmöglicher Starttermin: HS KW 38; FS KW 8

17.02.2020

2. Abgabetermin:

Dauer einer BAA: max. 15/16 Kalenderwochen (Um zur ordentlichen Diplomierung im Sommer zugelassen zu werden, muss die Abgabe bis spätestens Freitag, eine Woche nach Semesterende, erfolgen)

Dauer eines WIPRO: max. 14/15 Kalenderwochen

05.06.2020

3. Studierende:

	Student/in 1:	Student/in 2:
Name, Vorname:	<u>Anna Magdalena Huber</u>	
Studiengang:	<u>BSCI</u>	
Mobile:		
E-Mail:	<u>annamagdalena.huber@stud.hslu.ch</u>	
Projekt mit Arbeitgeber (bb-Studierende)	<input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nein	<input type="checkbox"/> Ja <input type="checkbox"/> Nein

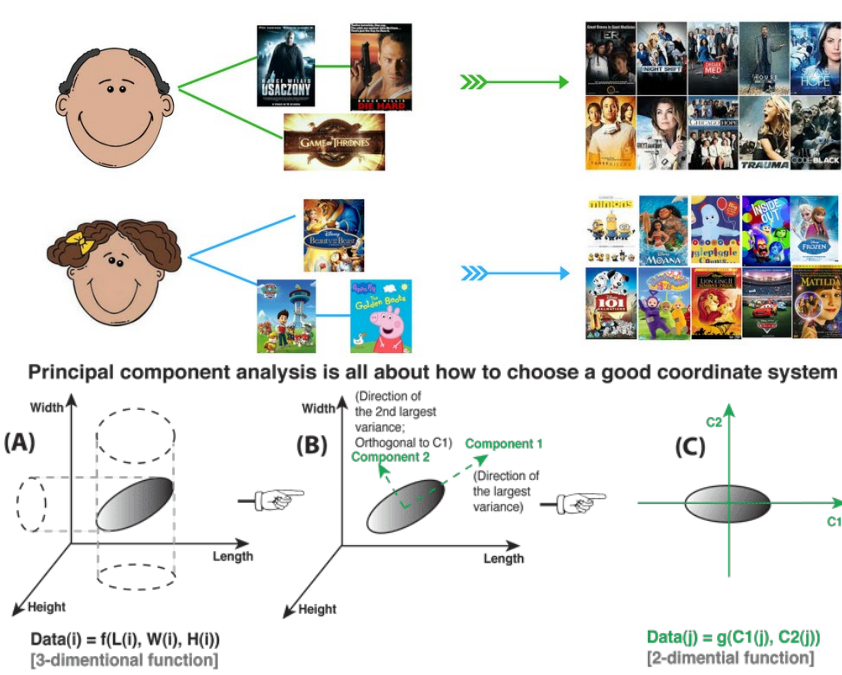
4. Auftraggeber/in:

Firma:	<u>Hochschule Luzern Informatik</u>
Ansprechperson:	<u>René Meier, Tahir Majeed</u>
Funktion:	
Strasse:	
PLZ / Ort:	
Telefon:	
Email:	<u>rene.meier@hslu.ch, tahir.majeed@hslu.ch</u>
Website:	

5. Betreuende/r Dozent/in:

René Meier und Tahir Majeed

6. Aufgabenstellung

<p>Titel:</p>	<p>Computing user and item neighbors in the PCA space for Movie Recommender Systems</p>
<p>Ausgangslage und Problemstellung:</p>	<p>This bachelor project is in the domain of Machine Learning and Recommender Systems. Generally, user-user and item-item neighborhoods are computed in the feature space (using e.g. ratings data). The user/items that are neighbors of each other are then used in the computation of recommendations. Pearson Correlation or Cosine Similarity is used in the feature space to compute the distances. Afterwards, the Top-N (usu. 30-50) nearest/closest neighbors are identified. The idea of this project is to compute the user and item distances in the Principal Component Analysis (PCA) space. Using the distances in the PCA space users/items neighbors are then identified in the PCA space using e.g. Mahalanobis Distance.</p>  <p>Principal component analysis is all about how to choose a good coordinate system</p> <p>(A) $\text{Data}(i) = f(L(i), W(i), H(i))$ [3-dimensional function]</p> <p>(B) $\text{Data}(i) = f(C1(i), C2(i))$ [2-dimensional function]</p> <p>(C) $\text{Data}(j) = g(C1(j), C2(j))$ [2-dimensional function]</p>
<p>Ziel der Arbeit und erwartete Resultate:</p>	<p>The goals of this project include:</p> <ul style="list-style-type: none"> - A state-of-the-art review into the relevant techniques (Feature Space, Cosine Similarity, PCA, Mahalanobis Distance) - Design of a software architecture for calculating user-user (first priority) and item-item (second priority) top-n neighbours (30-50) in Feature Space and PCA - Implementation of the architecture to calculate similarities in both spaces - Comparison of the top-n neighbours calculated in both spaces based on the data set provided
<p>Gewünschte Methoden, Vorgehen:</p>	<ul style="list-style-type: none"> • It is required that the students should learn how to compute PCA of the given data and get an understanding of the transformation between the feature space (original space) in comparison to the PCA space. • Get an understanding of the Mahalanobis distance. <ul style="list-style-type: none"> - item-to-item distances should be computed for all items - user-to-user distances should be computed for all users • Get an understanding between the distances in the original space and the PCA space. • Get an understanding of different evaluation measures to comparing two datasets.

	<ul style="list-style-type: none"> • Students are not required to implement the PCA approach. They can use any freely available implementation of computing PCA. • Students may use JAVA or Python for the task. • Apply PCA to the movies dataset. • Computed distances between item-item and user-user in the PCA space. • After computing the PCA, Mahalanobis distance (distances between items-items and distances between users-users) between the data points should be computed. <ul style="list-style-type: none"> ○ using all the eigenvectors. ○ using eigenvectors that explains 90% of the data variance. • Compute item-item and user-user distances in the original space. <ul style="list-style-type: none"> ○ use either Persons Correlation or Cosine Similarity. • Compare the Top-N (N could be 30, 50 and 100) neighbors using some suitable measures provide how much similar the two-neighborhood list (neighborhood in the PCA space and neighborhood in the original space) are. • Students should identify a suitable metric to compare two lists. • Dataset is huge; therefore, students will have to deal with the computational complexity of the algorithm and find approach to make it run faster. <ul style="list-style-type: none"> ○ 20000263 ratings ○ 27278 movies (items) ○ 138493 users (users) • Explore approaches to performing the computations using parallel processing libraries <ul style="list-style-type: none"> ○ TensorFlow • Results should be written to MySQL database. • Movies dataset will be provided.
Kreativität, Varianten, Innovation*	<p>Students can be creative in their solution for performing the computations in parallel to gain computational speed up.</p> <p>They can also be creative in performing comparison between the two list and choosing the right evaluation metrics.</p>
Schlagwörter:	<p>movies recommender system</p> <p>PCA</p> <p>cosine similarity</p> <p>Mahalanobis distance</p>
Wirtschaftsprojekt oder Bachelorarbeit:	<p><input checked="" type="checkbox"/> Wirtschaftsprojekt: 180 Stunden pro Studierender</p> <p><input type="checkbox"/> Bachelorarbeit: 360 Stunden</p>

* Bitte heben Sie in diesem Punkt hervor, inwiefern Ihre Projektidee **über kreativen Spielraum** verfügt. Dabei sind folgende Kriterien relevant: Die Idee erlaubt den Studierenden eigene Ideen zu entwickeln und Varianten zu erarbeiten, ist ausserhalb vom Tagesgeschäft angesiedelt, beinhaltet Neuland/Innovation und ist nicht durch Produkte & Tools getrieben.

Bitte kreuzen Sie eine Projektart und die zutreffenden Schwerpunkte an.

Projektarten:

- ☐ Einsatz von Standardsoftware und Services
- ☐ Software- und Produkt-Entwicklung
- ☒ Innovationsprojekte (Projekte mit Erkenntnisgewinn, Forschungsprojekte)
- ☐ IT-Infrastrukturentwicklung
- ☐ Strukturierte Analyse und Konzeption von Systemen und Abläufen

Schwerpunkte:

- ☒ Artificial Intelligence & Machine Learning
- ☐ Business Process Modelling
- ☐ Data Science
- ☐ Hardwarenahe Software-Erstellung
- ☐ Human Computer Interaction Design
- ☐ ICT Business Solutions
- ☐ ICT Infrastrukturen
- ☐ Internet of Things
- ☐ Mobile Systems
- ☐ Security/Privacy
- ☒ Software-Erstellung
- ☐ Visual Computing (Grafik, Bildverarbeitung, Vision, VR, AR)
- ☐ _____
- ☐ _____

7. Zeiteinteilung

Vorschlag für die Zeiteinteilung pro Person

WIPRO:

pro Woche: ca. 12h
für Modulendprüfung: ca. 10h
Total: 180 h

BAA:

pro Woche: ca. 20h
Schlusswoche: ca. 50h
Für Modulendprüfung: ca. 10h
Total: 360 h

8. Rechtliche Grundlagen und Reglemente

Folgende Rechtsgrundlagen und Reglemente sind für die Wirtschaftsprojekte und Bachelorarbeiten an der Hochschule Luzern – Informatik massgebend:

- Studienordnung für die Ausbildung an der Hochschule Luzern, FH Zentralschweiz ([Link](#))
- Studienreglement für die Bachelor-Ausbildung an der Hochschule Luzern - Informatik ([Link](#))

9. Bestätigung

Mit der Kenntnisnahme der Aufgabenstellung bestätigen Student/in und Auftraggeber/in, dass

- Sie mit der Aufgabenstellung einverstanden sind.
- die Auftraggeberin/der Auftraggeber damit einverstanden ist, dass die Hochschule Luzern – Informatik für die Organisation einer Bachelorarbeit von ihr/ihm einen Kostenbeitrag von CHF 1'000.00 (inkl. MwSt.) pro Student/in erhebt. Dies gilt nicht für Arbeiten, welche berufsbegleitend Studierende in Verbindung mit ihrem Arbeitgeber/ihrer Arbeitgeberin machen und für HSLU interne Auftraggeber/innen. Für die Wirtschaftsprojekte wird kein Kostenbeitrag verrechnet.
- Betreuungspersonen und Experten uneingeschränkter Einblick in die Arbeit erhalten. Auch anlässlich von Präsentationen und Marketingaktivitäten kann die Arbeit der Öffentlichkeit gezeigt werden. Eine Zusammenfassung der Arbeit wird in jedem Fall veröffentlicht. Falls das Thema vertraulich behandelt werden soll, muss der Aufgabenstellung eine entsprechende Vertraulichkeitserklärung beiliegen.

Datum: 19.02.2020

Die definitive Aufgabenstellung (pdf-Format) bitte per E-Mail an die Transferstelle senden,

zwingend in Kopie an alle involvierten Parteien.

Anlaufstelle für alle Informationen im Zusammenhang mit studentischen Arbeiten sowie für
Entgegennahme von Projektideen & Aufgabenstellungen:

Hochschule Luzern - Informatik
Transfer Services
Suurstoffi 1
6343 Rotkreuz
T: 041 228 24 66
E: transfer.informatik@hslu.ch