PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: DANIEL C. KLOKE
SHEPPARD, MULLIN, RICHTER & HAMPTON LLP
379 LYTTON AVENUE
PALO ALTO, CA 94301

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	24 SEP 2018
Applicant's or agent's file reference: 20WH-259182	FOR FURTHER ACTION — See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2018/040151	28 June 2018	28 June 2017

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - G06F 17/30; G06F 3/048; G06K 9/62 (2018.01)
CPC - G06F 17/30601; G06F 17/30604; G06F 17/30958; G06F 19/00; G06N 5/025 (2018.08)

Applicant AYASDI, INC.

1. This opinion contains indications relating to the following items:

- [x] Box No. I Basis of the opinion
- [ ] Box No. II Priority
- [ ] Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- [ ] Box No. IV Lack of unity of invention
- [x] Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- [ ] Box No. VI Certain documents cited
- [ ] Box No. VII Certain defects in the international application
- [ ] Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	30 August 2018	Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US2018/040151

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into ____________ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. [ ] This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. [ ] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:
   a. [ ] forming part of the international application as filed:
      - [ ] in the form of an Annex C/ST.25 text file.
      - [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:
      - [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
      - [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2018/040151

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-20	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-20	NO
Industrial applicability (IA)	Claims	1-20	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-20 lack an inventive step under PCT Article 33(3) as being obvious over Ayasdi, Inc. (hereinafter Ayasdi) in view of Wright et al. (hereinafter Wright).

Regarding claim 1, Ayasdi discloses a non-transitory computer readable medium including executable instructions (a non-transitory computer readable medium includes executable instructions, para 0008), the instructions being executable by a processor to perform a method (The instructions are executable by a processor to perform a method, para 0008), the method comprising:
receiving a network of a plurality of nodes and a plurality of edges (receiving a network of a plurality of nodes and a plurality of edges, para 0008), each of the nodes of the plurality of nodes comprising members representative of at least one subset of initial data points (each of the nodes of the plurality of nodes comprising members representative of at least one subset of training data points [i.e. initial data points], para 0008), each of the edges of the plurality of edges connecting nodes that share at least one data point of the initial data points (each of the edges of the plurality of edges connecting nodes that share at least one data point of the training data points [i.e. initial data points], para 0008), the initial data points including rows and columns (the training data set [i.e. initial data points] including rows and columns, para 0008), each row defining a data point of an initial data set and each column defining a feature (each row defining a data point of the training data set [i.e. initial data set] and each column defining a feature, para 0008), the initial data set including an initial number of columns (the training data set [i.e. initial data set] including an initial number of columns, para 0008), each column including values associated with a feature of a plurality of features (each column including values associated with a feature of a plurality of features, para 0008);
selecting a subset of the data points to create a set of selected data points (the data module 2802 may select a training data set from the initial data set. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology [i.e. selecting a subset of the data points to create a set of selected data points], para 0435), the selection being based on each node of the plurality of nodes (the interactive visualization allows the user to select nodes comprising data that has been clustered, para 0085, Once the objects are selected, a selection information box 912 may display some information regarding the selection, selection information box 912 indicates the number of nodes selected and the total points (e.g., data points or elements) of the selected nodes [i.e. the selection being based on nodes], para 0183), whereby if there is only one data point that is a member of a particular node, then the one data point is selected to be a member of the set of selected data points (Once the objects are selected, a selection information box 912 may display some information regarding the selection, selection information box 912 indicates the number of nodes selected and the total points (e.g., data points or elements) of the selected nodes, para 0183, The interactive visualization 900 may comprise any number of menu items. The "Selection" menu may allow the following functions: Select singletons (select nodes which are not connected to other nodes), para 0190-0191 A forest F is 'atomic' if every leaf in F is a singleton (e.g., a set with one member) [i.e. if there is only one data point that is a member of a particular node, then the one data point is selected to be a member of the set of selected data points], para 0296);
for each selected data point of the set of selected data points, determining a predetermined number of other data points of the set of selected data points that are closest in distance to that particular selected data point (One example of a hierarchical clustering technique, KNN on a finite metric space X is to compute the K nearest neighbors for each point of a network graph (e.g., a visualized or non-visualized graph that includes nodes that may be coupled to one or more other nodes of the graph) with, for example, K=50 [i.e. a predetermined number]. The partition generation module 2104 may start with INITIAL( ) being Singletons(X). Then at each step for 1<=k<=50, the partition generation module 2104 may connect x to y provided x and y are in the symmetric k nearest neighbors of one another, para 0323, The set S are the projections of the original data points into the reference space (e.g., a function such as a gausian density function is applied on the received data points to project to the reference space). The autogroup module 2002 may operate on a weighted graph built from this projection of the data into the reference space, for a fixed positive integer k, construct a graph G on the set S by connecting each point a in S to every point b in S if b is one of a's k-nearest neighbors and a is one of b's k-nearest neighbors (i.e. they are symmetric k-nearest neighbors of each other). In some testing, k=20 [i.e. a predetermined number] produces good results. The edges of the graph may be weighted by the distance between the edge's endpoints in the embedded reference space distance. This autogrouping embodiment may utilize a hierarchical single-linkage clusterer that uses distance between points in the reference space [i.e. for each selected data point of the set of selected data points, determining a predetermined number of other data points of the set of selected data points that are closest in distance to that particular selected data point], para 0410), the distance being determined based on a metric function between a vector of each data point (The analysis module 320 may perform analysis using the metric as a part of a distance function. The distance function can be expressed by a formula, a distance matrix, or other routine which computes it, para 0142, the set S is a set data together with a metric which defines a distance between any two points in the set S [i.e. the distance being determined based on a metric function between a vector of each data point], para 0413);
grouping the selected data points into a plurality of groups based, at least in part, on the predetermined number of other data points of the set of selected data points that are closest in distance (grouping the data points of the training data set into a plurality of groups, para 0008, One example of a hierarchical clustering technique, KNN on a finite metric space X is to compute the K nearest neighbors for each

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US2018/040151

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

point of a network graph (e.g., a visualized or non-visualized graph that includes nodes that may be coupled to one or more other nodes of the graph) with, for example, K=50, para 0323, The autogroup module 2002 may operate on a weighted graph built from this projection of the data into the reference space. For example, for a fixed positive integer k, construct a graph G on the set S by connecting each point a in S to every point b in S if b is one of a's k-nearest neighbors and a is one of b's k-nearest neighbors (i.e. they are symmetric k-nearest neighbors of each other). In some testing, k=20 produces good results. The edges of the graph may be weighted by the distance between the edge's endpoints in the embedded reference space distance. This autogrouping embodiment may utilize a hierarchical single-linkage clusterer that uses distance between points in the reference space, para 0410, a partition P of the data points in the original space. For a fixed positive integer "j", we can expand each subset "a" of P by adding all the j-nearest neighbors of the elements in the subset "a" [i.e. grouping the selected data points into a plurality of groups based, at least in part, on the predetermined number of other data points of the set of selected data points that are closest in distance], para 0418), each group of the plurality of groups including a different subset of data points (each group of the plurality of groups including a different subset of data points of the training data set, para 0008); and
providing a list of selected data points and the plurality of groups (the data control module 2114 generates a report indicating the selected and/or generated partition from the partition selection module 2112. The report may include, for example, data sets, partitions, subsets, elements, data set identifiers, partition identifiers, subset identifiers, element identifiers, and/or the like. In some embodiments, the report may include a graph (e.g., see FIG. 19) with an indication of selected nodes whose member(s) include data of the selected and/or generated partition from the partition selection module 2112, para 0361, an example report 2400 of an autogrouped graph of data points that depicts the grouped data [i.e. providing a list of selected data points and the plurality of groups], para 0400).
Ayasdi does not specifically disclose that if there are two or more data points that are a member of the particular node, then proportional number of data points relative to all data points that are members of that particular node are selected to be members of the set of selected data points.
However, Wright is in the field of data visualization of large data sets through data reduction techniques (para 0002) and discloses that if there are two or more data points that are a member of the particular node, then proportional number of data points relative to all data points that are members of that particular node are selected to be members of the set of selected data points (due to the massive amount of data, the node-and-link representation often does not achieve the density possible with a bitmap, consider that in a two dimensional digital image it is difficult to visually represent more distinct data points than the number of pixels used to draw that image, para 0004, a data reduction module for reducing the original data set to produce a reduced data set having a number of reduced data points less than the number of original data points. The number of reduced data points is based on a received query parameter, a data resizing module for dynamically resizing the received reduced data set to produce a resized data set suitable for use in generating a display of pixels appropriate to the number of available pixels. The data resizing module is configured for summing or otherwise combining the individual data values of selected adjacent ones of the reduced data points in the reduced data set and assigning the summed value to a respective data value of a resized data point in the resized data set [i.e. proportional number of data points relative to all data points are selected], para 0013, This function can be logarithmic in the number of data chunks 482, which in turn can be proportional to the square root of the number of data points, para 0117). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify Ayasdi with the teachings of Wright for the purpose of reducing the original data set to produce a reduced data set having a number of reduced data points less than the number of original data points and dynamically resizing the received reduced data set to produce a resized data set suitable for use in generating a display of pixels appropriate to the number of available pixels (Wright abstract).

Regarding claim 2, Ayasdi further discloses the method further comprising:
creating a first transformation data set, the first transformation data set including the selected data points as well as a plurality of feature subsets, each of the plurality of feature subsets being associated with at least one group of the plurality of groups, values of a particular data point for a particular feature subset for a particular group being based on values of the particular data point in the selected data points if the particular data point is a member of the particular group (creating a first transformation data set, the first transformation data set including the training data set as well as a plurality of feature subsets, each of the plurality of feature subsets being associated with at least one group of the plurality of groups, values of a particular data point for a particular feature subset for a particular group being based on values of the particular data point in the training data set if the particular data point is a member of the particular group, para 0008, the data module 2802 may select a training data set from the initial data set [i.e. the selected data points]. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology, para 0435); and
applying a machine learning model to the first transformation data set to generate a prediction model (applying a machine learning model to the first transformation data set to generate a prediction model, para 0008).

Regarding claim 3, Ayasdi further discloses the method further comprising:
creating a second transformation data set, the second transformation data set including the analysis data set as well as the plurality of feature subsets, each of the plurality of feature subsets being associated with the at least one group of the plurality of groups, values of a particular data point of the analysis data set for a particular feature subset for a particular group being based on values of the particular data point in the analysis data set if the particular data point is a member of the particular group (creating a second transformation data set, the second transformation data set including the analysis data set as well as the plurality of feature subsets, each of the plurality of feature subsets being associated with the at least one group of the plurality of groups, values of a particular data point of the analysis data set for a particular feature subset for a particular group being based on values of the particular data point in the analysis data set if the particular data point is a member of the particular group, para 0009);
applying the prediction model to the second transformation data set to generate predicted outcomes (applying the prediction model to the second transformation data set to generate predicted outcomes, para 0009); and
generating a report indicating one or more of the predicted outcomes (generating a report indicating one or more of the predicted outcomes, para 0009).

Regarding claim 4, Ayasdi further discloses the method further comprising comparing the predicted outcomes to known outcomes to assess the quality of the prediction model (comparing the predicted outcomes to known outcomes to assess the quality of the prediction model, para 0010).

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2018/040151

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 5, Ayasdi further discloses that the network of the plurality of nodes and the plurality of edges are a result of topological data analysis applied to the initial data set (The network of the plurality of nodes and the plurality of edges may be a result of topological data analysis applied to the training data set, para 0011).

Regarding claim 6, Ayasdi further discloses that the network of the plurality of nodes and the plurality of edges are generated by:
generating a reference space (generating a reference space, para 0011);
mapping the data points of the training data into the reference space using at least one filter (mapping the data points of the training data into the reference space using at least one filter, para 0011);
generating a cover based on a resolution (generating a cover based on a resolution, para 0011);
clustering data in the cover based on a metric and data points of the training data set (clustering data in the cover based on a metric and data points of the training data set, para 0011);
identifying nodes based on the clustered data (identifying nodes based on the clustered data, para 0011); and
identifying edges between nodes if nodes share member data points from the training data set (identifying edges between nodes if nodes share member data points from the training data set, para 0011).

Regarding claim 7, Ayasdi further discloses that values of a particular data point for a particular feature subset for a particular group are zero if the particular data point of the training data set is not a member of the particular group (values of a particular data point for a particular feature subset for a particular group are zero if the particular data point of the training data set is not a member of the particular group, para 0012).

Regarding claim 8, Ayasdi further discloses that values of a particular data point for a particular feature subset for a particular group are null if the particular data point of the training data set is not a member of the particular group (the values of a particular data point for a particular feature subset for a particular group are null if the particular data point of the training data set is not a member of the particular group, para 0012).

Regarding claim 9, Ayasdi further discloses that the values of a particular data point for a particular feature subset for a particular group of which the particular data point is a member are weighted (The values of a particular data point for a particular feature subset for a particular group of which the particular data point may be a member are weighted, para 0012).

Regarding claim 10, Ayasdi further discloses that weighting of the values for the particular data point at least partially depend on how many the plurality of groups the particular data point is a member of (Weighting of the values for the particular data point may at least partially depend on how many the plurality of groups the particular data point is a member of, para 0012).

Regarding claim 11, Ayasdi further discloses that the machine learning model is selected from a group consisting of a linear regression machine learning model, a polynomial regression machine learning model, a logistic regression machine learning model, and a random forest machine learning model (The machine learning model may be selected from a group consisting of a linear regression machine learning model, a polynomial regression machine learning model, a logistic regression machine learning model, and a random forest machine learning model, para 0013).

Regarding claim 12, Ayasdi further discloses further comprising:
generating a reference space (generating a reference space, para 0011);
mapping the selected data points data into the reference space using at least one filter function (mapping the data points of the training data into the reference space using at least one filter, para 0011, the data module 2802 may select a training data set from the initial data set [i.e. the selected data points]. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology, para 0435);
generating a cover based on a resolution (generating a cover based on a resolution, para 0011);
clustering data in the cover based on a metric function and selected data points (clustering data in the cover based on a metric and data points of the training data set, para 0011, the data module 2802 may select a training data set from the initial data set [i.e. the selected data points]. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology, para 0435);
identifying new nodes based on the clustered data (identifying nodes based on the clustered data, para 0011, a new patient may be localized on the map visualization. With the map visualization for subtypes of a particular disease and a new patient diagnosed with the disease, point(s) may be located among the data points used in computing the map visualization (e.g., nearest neighbor) which is closest to the new patient point. The new patient may be labeled with nodes in the map visualization containing the closest neighbor. These nodes may be highlighted to give a physician the location of the new patient among the patients in the reference data set, para 0224);
identifying new edges between new nodes if nodes share member selected data points (identifying edges between nodes if nodes share member data points from the training data set, para 0011, the new patient distance module determines distances between the biological data of each patient of the cancer map visualization 1600 and the new biological data from the new patient. For example, the previous biological data that was utilized in the generation of the cancer map visualization 1600 may be stored in mapped data structures. Distances may be determined between the new biological data of the new patient and each of the previous patient's biological data in the mapped data structure, para 0252); and
providing a display of the selected data points and node membership (The output may be, for example, a visualization (e.g., a display of connected nodes or "network") or simplicial complex. One specific combinatorial formulation in one embodiment may be that the vertices form a finite set, and then the additional structure may be a collection of edges (unordered pairs of vertices) which are pictured as connections in this network, para 0071, The visualization engine 322 may also display information regarding the selection (e.g., by displaying a selection information window 912), para 0214).

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2018/040151

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 13, Ayasdi discloses a method comprising:
receiving a network of a plurality of nodes and a plurality of edges (receiving a network of a plurality of nodes and a plurality of edges, para 0008), each of the nodes of the plurality of nodes comprising members representative of at least one subset of initial data points (each of the nodes of the plurality of nodes comprising members representative of at least one subset of training data points [i.e. initial data points], para 0008), each of the edges of the plurality of edges connecting nodes that share at least one data point of the initial data points (each of the edges of the plurality of edges connecting nodes that share at least one data point of the training data points [i.e. initial data points], para 0008), the initial data points including rows and columns (the training data set [i.e. initial data points] including rows and columns, para 0008), each row defining a data point of an initial data set and each column defining a feature (each row defining a data point of the training data set [i.e. initial data set] and each column defining a feature, para 0008), the initial data set including an initial number of columns (the training data set [i.e. initial data set] including an initial number of columns, para 0008), each column including values associated with a feature of a plurality of features (each column including values associated with a feature of a plurality of features, para 0008);
selecting a subset of the data points to create a set of selected data points (the data module 2802 may select a training data set from the initial data set. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology [i.e. selecting a subset of the data points to create a set of selected data points], para 0435), the selection being based on each node of the plurality of nodes (the interactive visualization allows the user to select nodes comprising data that has been clustered, para 0085, Once the objects are selected, a selection information box 912 may display some information regarding the selection, selection information box 912 indicates the number of nodes selected and the total points (e.g., data points or elements) of the selected nodes [i.e. the selection being based on nodes], para 0183), whereby if there is only one data point that is a member of a particular node, then the one data point is selected to be a member of the set of selected data points (Once the objects are selected, a selection information box 912 may display some information regarding the selection, selection information box 912 indicates the number of nodes selected and the total points (e.g., data points or elements) of the selected nodes, para 0183, The interactive visualization 900 may comprise any number of menu items. The "Selection" menu may allow the following functions: Select singletons (select nodes which are not connected to other nodes), para 0190-0191 A forest F is 'atomic' if every leaf in F is a singleton (e.g., a set with one member) [i.e. if there is only one data point that is a member of a particular node, then the one data point is selected to be a member of the set of selected data points], para 0296);
for each selected data point of the set of selected data points, determining a predetermined number of other data points of the set of selected data points that are closest in distance to that particular selected data point (One example of a hierarchical clustering technique, KNN on a finite metric space X is to compute the K nearest neighbors for each point of a network graph (e.g., a visualized or non-visualized graph that includes nodes that may be coupled to one or more other nodes of the graph) with, for example, K=50 [i.e. a predetermined number]. The partition generation module 2104 may start with INITIAL( ) being Singletons(X). Then at each step for 1<=k<=50, the partition generation module 2104 may connect x to y provided x and y are in the symmetric k nearest neighbors of one another, para 0323, The set S are the projections of the original data points into the reference space (e.g., a function such as a gausian density function is applied on the received data points to project to the reference space). The autogroup module 2002 may operate on a weighted graph built from this projection of the data into the reference space, for a fixed positive integer k, construct a graph G on the set S by connecting each point a in S to every point b in S if b is one of a's k-nearest neighbors and a is one of b's k-nearest neighbors (i.e. they are symmetric k-nearest neighbors of each other). In some testing, k=20 [i.e. a predetermined number] produces good results. The edges of the graph may be weighted by the distance between the edge's endpoints in the embedded reference space distance. This autogrouping embodiment may utilize a hierarchical single-linkage clusterer that uses distance between points in the reference space [i.e. for each selected data point of the set of selected data points, determining a predetermined number of other data points of the set of selected data points that are closest in distance to that particular selected data point], para 0410), the distance being determined based on a metric function between a vector of each data point (The analysis module 320 may perform analysis using the metric as a part of a distance function. The distance function can be expressed by a formula, a distance matrix, or other routine which computes it, para 0142, the set S is a set data together with a metric which defines a distance between any two points in the set S [i.e. the distance being determined based on a metric function between a vector of each data point], para 0413);
grouping the selected data points into a plurality of groups based, at least in part, on the predetermined number of other data points of the set of selected data points that are closest in distance (grouping the data points of the training data set into a plurality of groups, para 0008, One example of a hierarchical clustering technique, KNN on a finite metric space X is to compute the K nearest neighbors for each point of a network graph (e.g., a visualized or non-visualized graph that includes nodes that may be coupled to one or more other nodes of the graph) with, for example, K=50, para 0323, The autogroup module 2002 may operate on a weighted graph built from this projection of the data into the reference space. For example, for a fixed positive integer k, construct a graph G on the set S by connecting each point a in S to every point b in S if b is one of a's k-nearest neighbors and a is one of b's k-nearest neighbors (i.e. they are symmetric k-nearest neighbors of each other). In some testing, k=20 produces good results. The edges of the graph may be weighted by the distance between the edge's endpoints in the embedded reference space distance. This autogrouping embodiment may utilize a hierarchical single-linkage clusterer that uses distance between points in the reference space, para 0410, a partition P of the data points in the original space. For a fixed positive integer "j", we can expand each subset "a" of P by adding all the j-nearest neighbors of the elements in the subset "a" [i.e. grouping the selected data points into a plurality of groups based, at least in part, on the predetermined number of other data points of the set of selected data points that are closest in distance], para 0418), each group of the plurality of groups including a different subset of data points (each group of the plurality of groups including a different subset of data points of the training data set, para 0008); and
providing a list of selected data points and the plurality of groups (the data control module 2114 generates a report indicating the selected and/or generated partition from the partition selection module 2112. The report may include, for example, data sets, partitions, subsets, elements, data set identifiers, partition identifiers, subset identifiers, element identifiers, and/or the like. In some embodiments, the report may include a graph (e.g., see FIG. 19) with an indication of selected nodes whose member(s) include data of the selected and/or generated partition from the partition selection module 2112, para 0361, an example report 2400 of an autogrouped graph of data points that depicts the grouped data [i.e. providing a list of selected data points and the plurality of groups], para 0400).
Ayasdi does not specifically disclose that if there are two or more data points that are a member of the particular node, then proportional number of data points relative to all data points that are members of that particular node are selected to be members of the set of selected data points.
However, Wright discloses that if there are two or more data points that are a member of the particular node, then proportional number of

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2018/040151

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

data points relative to all data points that are members of that particular node are selected to be members of the set of selected data points (due to the massive amount of data, the node-and-link representation often does not achieve the density possible with a bitmap, consider that in a two dimensional digital image it is difficult to visually represent more distinct data points than the number of pixels used to draw that image, para 0004, a data reduction module for reducing the original data set to produce a reduced data set having a number of reduced data points less than the number of original data points. The number of reduced data points is based on a received query parameter, a data resizing module for dynamically resizing the received reduced data set to produce a resized data set suitable for use in generating a display of pixels appropriate to the number of available pixels. The data resizing module is configured for summing or otherwise combining the individual data values of selected adjacent ones of the reduced data points in the reduced data set and assigning the summed value to a respective data value of a resized data point in the resized data set [i.e. proportional number of data points relative to all data points are selected], para 0013, This function can be logarithmic in the number of data chunks 482, which in turn can be proportional to the square root of the number of data points, para 0117). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify Ayasdi with the teachings of Wright for the purpose of reducing the original data set to produce a reduced data set having a number of reduced data points less than the number of original data points and dynamically resizing the received reduced data set to produce a resized data set suitable for use in generating a display of pixels appropriate to the number of available pixels (Wright abstract).

Regarding claim 14, Ayasdi further discloses further comprising:
creating a first transformation data set, the first transformation data set including the selected data points as well as a plurality of feature subsets, each of the plurality of feature subsets being associated with at least one group of the plurality of groups, values of a particular data point for a particular feature subset for a particular group being based on values of the particular data point in the selected data points if the particular data point is a member of the particular group (creating a first transformation data set, the first transformation data set including the training data set as well as a plurality of feature subsets, each of the plurality of feature subsets being associated with at least one group of the plurality of groups, values of a particular data point for a particular feature subset for a particular group being based on values of the particular data point in the training data set if the particular data point is a member of the particular group, para 0008, the data module 2802 may select a training data set from the initial data set [i.e. the selected data points]. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology, para 0435); and
applying a machine learning model to the first transformation data set to generate a prediction model (applying a machine learning model to the first transformation data set to generate a prediction model, para 0008).

Regarding claim 15, Ayasdi further discloses further comprising:
creating a first transformation data set, the first transformation data set including the selected data points as well as a plurality of feature subsets, each of the plurality of feature subsets being associated with at least one group of the plurality of groups, values of a particular data point for a particular feature subset for a particular group being based on values of the particular data point in the selected data points if the particular data point is a member of the particular group (creating a first transformation data set, the first transformation data set including the training data set as well as a plurality of feature subsets, each of the plurality of feature subsets being associated with at least one group of the plurality of groups, values of a particular data point for a particular feature subset for a particular group being based on values of the particular data point in the training data set if the particular data point is a member of the particular group, para 0008, the data module 2802 may select a training data set from the initial data set [i.e. the selected data points]. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology, para 0435); and
applying a machine learning model to the first transformation data set to generate a prediction model (applying a machine learning model to the first transformation data set to generate a prediction model, para 0008).

Regarding claim 16, Ayasdi further discloses further comprising:
creating a second transformation data set, the second transformation data set including the analysis data set as well as the plurality of feature subsets, each of the plurality of feature subsets being associated with the at least one group of the plurality of groups, values of a particular data point of the analysis data set for a particular feature subset for a particular group being based on values of the particular data point in the analysis data set if the particular data point is a member of the particular group (creating a second transformation data set, the second transformation data set including the analysis data set as well as the plurality of feature subsets, each of the plurality of feature subsets being associated with the at least one group of the plurality of groups, values of a particular data point of the analysis data set for a particular feature subset for a particular group being based on values of the particular data point in the analysis data set if the particular data point is a member of the particular group, para 0009);
applying the prediction model to the second transformation data set to generate predicted outcomes (applying the prediction model to the second transformation data set to generate predicted outcomes, para 0009); and
generating a report indicating one or more of the predicted outcomes (generating a report indicating one or more of the predicted outcomes, para 0009).

Regarding claim 17, Ayasdi further discloses further comprising comparing the predicted outcomes to known outcomes to assess the quality of the prediction model (comparing the predicted outcomes to known outcomes to assess the quality of the prediction model, para 0010).

Regarding claim 18, Ayasdi further discloses that the network of the plurality of nodes and the plurality of edges are a result of topological data analysis applied to the initial data set (The network of the plurality of nodes and the plurality of edges may be a result of topological data analysis applied to the training data set, para 0011).

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US2018/040151

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 19, Ayasdi further discloses that the network of the plurality of nodes and the plurality of edges are generated by:
receiving the training data set (receiving the training data set, para 0011);
generating a reference space (generating a reference space, para 0011);
mapping the data points of the training data into the reference space using at least one filter (mapping the data points of the training data into the reference space using at least one filter, para 0011);
generating a cover based on a resolution (generating a cover based on a resolution, para 0011);
clustering data in the cover based on a metric and data points of the training data set (clustering data in the cover based on a metric and data points of the training data set, para 0011);
identifying nodes based on the clustered data (identifying nodes based on the clustered data, para 0011); and
identifying edges between nodes if nodes share member data points from the training data set (identifying edges between nodes if nodes share member data points from the training data set, para 0011).

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No. PCT/US2018/040151

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

Regarding claim 20, Ayasdi discloses a system comprising:
a processor (The instructions are executable by a processor to perform a method, para 0008); and
a memory, the memory comprising instructions executable by the processor to perform the steps of (a non-transitory computer readable medium includes executable instructions. The instructions are executable by a processor to perform a method, para 0008):
receiving a network of a plurality of nodes and a plurality of edges (receiving a network of a plurality of nodes and a plurality of edges, para 0008), each of the nodes of the plurality of nodes comprising members representative of at least one subset of initial data points (each of the nodes of the plurality of nodes comprising members representative of at least one subset of training data points [i.e. initial data points], para 0008), each of the edges of the plurality of edges connecting nodes that share at least one data point of the initial data points (each of the edges of the plurality of edges connecting nodes that share at least one data point of the training data points [i.e. initial data points], para 0008), the initial data points including rows and columns (the training data set [i.e. initial data points] including rows and columns, para 0008), each row defining a data point of an initial data set and each column defining a feature (each row defining a data point of the training data set [i.e. initial data set] and each column defining a feature, para 0008), the initial data set including an initial number of columns (the training data set [i.e. initial data set] including an initial number of columns, para 0008), each column including values associated with a feature of a plurality of features (each column including values associated with a feature of a plurality of features, para 0008);
selecting a subset of the data points to create a set of selected data points (the data module 2802 may select a training data set from the initial data set. The data module 2802 may select the training data set in any number of ways, the data module 2802 may select a subset of data points from the initial data set randomly or using any methodology [i.e. selecting a subset of the data points to create a set of selected data points], para 0435), the selection being based on each node of the plurality of nodes (the interactive visualization allows the user to select nodes comprising data that has been clustered, para 0085, Once the objects are selected, a selection information box 912 may display some information regarding the selection, selection information box 912 indicates the number of nodes selected and the total points (e.g., data points or elements) of the selected nodes [i.e. the selection being based on nodes], para 0183), whereby if there is only one data point that is a member of a particular node, then the one data point is selected to be a member of the set of selected data points (Once the objects are selected, a selection information box 912 may display some information regarding the selection, selection information box 912 indicates the number of nodes selected and the total points (e.g., data points or elements) of the selected nodes, para 0183, The interactive visualization 900 may comprise any number of menu items. The "Selection" menu may allow the following functions: Select singletons (select nodes which are not connected to other nodes), para 0190-0191 A forest F is 'atomic' if every leaf in F is a singleton (e.g., a set with one member) [i.e. if there is only one data point that is a member of a particular node, then the one data point is selected to be a member of the set of selected data points], para 0296);
for each selected data point of the set of selected data points, determining a predetermined number of other data points of the set of selected data points that are closest in distance to that particular selected data point (One example of a hierarchical clustering technique, KNN on a finite metric space X is to compute the K nearest neighbors for each point of a network graph (e.g., a visualized or non-visualized graph that includes nodes that may be coupled to one or more other nodes of the graph) with, for example, K=50 [i.e. a predetermined number]. The partition generation module 2104 may start with INITIAL( ) being Singletons(X). Then at each step for 1<=k<=50, the partition generation module 2104 may connect x to y provided x and y are in the symmetric k nearest neighbors of one another, para 0323, The set S are the projections of the original data points into the reference space (e.g., a function such as a gausian density function is applied on the received data points to project to the reference space). The autogroup module 2002 may operate on a weighted graph built from this projection of the data into the reference space, for a fixed positive integer k, construct a graph G on the set S by connecting each point a in S to every point b in S if b is one of a's k-nearest neighbors and a is one of b's k-nearest neighbors (i.e. they are symmetric k-nearest neighbors of each other). In some testing, k=20 [i.e. a predetermined number] produces good results. The edges of the graph may be weighted by the distance between the edge's endpoints in the embedded reference space distance. This autogrouping embodiment may utilize a hierarchical single-linkage clusterer that uses distance between points in the reference space [i.e. for each selected data point of the set of selected data points, determining a predetermined number of other data points of the set of selected data points that are closest in distance to that particular selected data point], para 0410), the distance being determined based on a metric function between a vector of each data point (The analysis module 320 may perform analysis using the metric as a part of a distance function. The distance function can be expressed by a formula, a distance matrix, or other routine which computes it, para 0142, the set S is a set data together with a metric which defines a distance between any two points in the set S [i.e. the distance being determined based on a metric function between a vector of each data point], para 0413);
grouping the selected data points into a plurality of groups based, at least in part, on the predetermined number of other data points of the set of selected data points that are closest in distance (grouping the data points of the training data set into a plurality of groups, para 0008, One example of a hierarchical clustering technique, KNN on a finite metric space X is to compute the K nearest neighbors for each point of a network graph (e.g., a visualized or non-visualized graph that includes nodes that may be coupled to one or more other nodes of the graph) with, for example, K=50, para 0323, The autogroup module 2002 may operate on a weighted graph built from this projection of the data into the reference space. For example, for a fixed positive integer k, construct a graph G on the set S by connecting each point a in S to every point b in S if b is one of a's k-nearest neighbors and a is one of b's k-nearest neighbors (i.e. they are symmetric k-nearest neighbors of each other). In some testing, k=20 produces good results. The edges of the graph may be weighted by the distance between the edge's endpoints in the embedded reference space distance. This autogrouping embodiment may utilize a hierarchical single-linkage clusterer that uses distance between points in the reference space, para 0410, a partition P of the data points in the original space. For a fixed positive integer "j", we can expand each subset "a" of P by adding all the j-nearest neighbors of the elements in the subset "a" [i.e. grouping the selected data points into a plurality of groups based, at least in part, on the predetermined number of other data points of the set of selected data points that are closest in distance], para 0418), each group of the plurality of groups including a different subset of data points (each group of the plurality of groups including a different subset of data points of the training data set, para 0008); and
providing a list of selected data points and the plurality of groups (the data control module 2114 generates a report indicating the selected and/or generated partition from the partition selection module 2112. The report may include, for example, data sets, partitions, subsets, elements, data set identifiers, partition identifiers, subset identifiers, element identifiers, and/or the like. In some embodiments, the report may include a graph (e.g., see FIG. 19) with an indication of selected nodes whose member(s) include data of the selected and/or generated partition from the partition selection module 2112, para 0361, an example report 2400 of an autogrouped graph of data points that depicts the grouped data [i.e. providing a list of selected data points and the plurality of groups], para 0400).
Ayasdi does not specifically disclose that if there are two or more data points that are a member of the particular node, then proportional

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No. PCT/US2018/040151

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

number of data points relative to all data points that are members of that particular node are selected to be members of the set of selected data points.
However, Wright discloses that if there are two or more data points that are a member of the particular node, then proportional number of data points relative to all data points that are members of that particular node are selected to be members of the set of selected data points (due to the massive amount of data, the node-and-link representation often does not achieve the density possible with a bitmap, consider that in a two dimensional digital image it is difficult to visually represent more distinct data points than the number of pixels used to draw that image, para 0004, a data reduction module for reducing the original data set to produce a reduced data set having a number of reduced data points less than the number of original data points. The number of reduced data points is based on a received query parameter, a data resizing module for dynamically resizing the received reduced data set to produce a resized data set suitable for use in generating a display of pixels appropriate to the number of available pixels. The data resizing module is configured for summing or otherwise combining the individual data values of selected adjacent ones of the reduced data points in the reduced data set and assigning the summed value to a respective data value of a resized data point in the resized data set [i.e. proportional number of data points relative to all data points are selected], para 0013, This function can be logarithmic in the number of data chunks 482, which in turn can be proportional to the square root of the number of data points, para 0117). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify Ayasdi with the teachings of Wright for the purpose of reducing the original data set to produce a reduced data set having a number of reduced data points less than the number of original data points and dynamically resizing the received reduced data set to produce a resized data set suitable for use in generating a display of pixels appropriate to the number of available pixels (Wright abstract).

Claims 1-20 meet the criteria set out in PCT Article 33(4), and thus have industrial applicability because the subject matter claimed can be made or used in industry.